The Role of Static and Dynamic Correlation in Molecular Electronic Structure, Spectroscopy, and Intermolecular Interactions

---

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Chemistry

at the

University of Canterbury, Christchurch, New Zealand

Andrew James Wallace

2016

Declaration

This thesis has not been previously submitted, in whole or in part, for any degree at this or any other university. The work is original and my own, carried out under the supervision of Dr Deborah Crittenden; where this is not so, credit has been duly given.

Andrew Wallace

Acknowledgements

I would like to thank my supervisor, Deborah Crittenden, for her endless support over the years on these projects, and the great deal of time she has dedicated to helping me work on them. My officemates also deserve credit for keeping me sane, or at least sharing my insanities, and providing much-needed conversation and entertainment over the years. I would like to thank the university for the Canterbury Scholarship that funded most of my PhD work. The support team for the NESI and Bluefern supercomputing facilities deserve a good deal of credit for the amount of time they have spent responding to my help-requests. Their provision of supercomputing time and resources has proven invaluable to the completion of this work.

Abstract

Electron correlation is a central problem in quantum chemistry. This thesis explores the role that static and dynamic correlation play in three different computational applications: The basis set dependence of energy calculations; the formation of a dicationic dimer; and the calculation of the spectra and ground states of metallophthalocyanines. The basis set dependence of the static correlation energy is established for a chemically diverse collection of atoms and molecules, and shown to converge exponentially with respect to basis set cardinality. For most practical purposes, a triple zeta basis set is sufficient to recover the static correlation energy with high accuracy. A new dicationic dimer system comprised of closed shell monomers is examined, to determine what forces are driving dimer formation. It is demonstrated that dynamic correlation energy overcomes the electrostatic repulsion between the cations, when supplemented by charge-balancing environmental effects, and therefore lead to dimer formation. Explaining the electronic spectra of the metallophthalocyanine molecules has posed a problem for spectroscopists for over 30 years. Static correlation is found to play a determining role in excited state energies due to the presence of a number of energetically-close π-orbitals. The experimental spectrum of zinc phthalocyanine is reproduced computationally with EOM-CCSD based methods giving the experimentally observed number of peaks all within 0.4 eV of observed experimental energies. Past DFT studies on the ground states of the Mn, Fe, and Co phthalocyanines have disagreed concerning the ground state orbital occupations of the metal ions for these systems. This thesis uses CASSCF-based methods to predict the ground states of these systems, which are determined to be Mn: $(d_{xy})^2(d_{xz},d_{yz})^{1,1}(d_{z^2})^1$, Co: $(d_{xy})^2(d_{xz},d_{yz})^{2,2}(d_{z^2})^1$. For Fe two possibilities for the ground state were found to be within 0.02 eV of each other in the highest-quality calculations: $(d_{xy})^2(d_{xz},d_{yz})^{1,1}(d_{z^2})^2$ and $(d_{xy})^2(d_{xz},d_{yz})^{2,1}(d_{z^2})^1$.

Deputy Vice-Chancellor's Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 2 is identical (save formatting) to the published journal article:
Andrew J. Wallace, Deborah L. Crittenden; Optimal Composition of Atomic Orbital Basis Sets for Recovering Static Correlation Energies*, J. Phys. Chem. A*, **2014**, *118* (11), pp 2138–2148

Please detail the nature and extent (%) of contribution by the candidate:

99% of the computational calculations were performed by the candidate (all calculations aside from the numerically-exact HF values in Table 8).

30% of the written content was provided by the candidate

Certification by Co-authors:
If there is more than one co-author then a single co-author can sign on behalf of all
The undersigned certifys that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: Deborah Crittenden Signature: D. Crittenden Date: 28/2/2016

Deputy Vice-Chancellor's Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 3 is identical (save formatting) to the published journal article:
Andrew J. Wallace, Chaminda D. Jayasinghe, Matthew I. J. Polson, Owen J. Curnow, and Deborah L. Crittenden*;* Cyclopropenium Cations Break the Rules of Attraction to Form Closely Bound Dimers, *J. Am. Chem. Soc.*, **2015**, *137* (49), pp 15528–15532

Please detail the nature and extent (%) of contribution by the candidate:

100% of the computational calculations were performed by the candidate.
40% of the written content was provided by the candidate

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all
The undersigned certifys that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: Deborah Crittenden Signature: D. Crittenden Date: 28/2/2016

Abbreviations and Acronyms

AO	Atomic orbital
cc-pVnZ	The correlation-consistent polarized valence *n*-zeta basis
CASSCF	Complete active space self-consistent field
CASPT2	Complete active space self-consistent field with second order Møller-Plesset perturbation theory energy correction
CIS	Configuration interaction with single excitations
CISDTQ	Configuration-interaction with single, double, triple, and quadruple excitations
CCSD	Coupled cluster with singles and doubles
CCSD(T)	Coupled cluster with singles, doubles and non-iterative triples
E_{dyn}	Dynamic correlation energy
E_{stat}	Static correlation energy
EOM-CC	Equation-of-motion coupled cluster
HF	Hartree Fock
HOMO	Highest occupied molecular orbital
LUMO	Lowest unoccupied molecular orbital
MO	Molecular orbital
MP2	Second order Møller-Plesset perturbation theory
MPc	Metallophthalocyanine
MRMP2	Multireference with second order Møller-Plesset perturbation theory energy correction
pc-*n*	The polarization consistent basis with *n* types of higher angular momentum functions
Pc	Phthalocyanine
SAPT	Symmetry-adapted perturbation theory
SOMO	Singly occupied molecular orbital
VOD	Valence-optimized doubles
ZnPc	Zinc Phthalocyanine

Table of contents

1. Introduction and Methods

Quantum chemical modeling

Quantum mechanical models provide the theoretical underpinnings for understanding chemical processes. Continuing improvements to these models have increasingly provided the ability to explain and interpret, or even predict, chemical processes. One of the most heavily used models in chemistry is molecular orbital theory, where electrons are modeled as occupying molecular orbitals (MOs) that are formed from linear combinations of the atomic orbitals in the molecule.[1] Energy level diagrams constructed using MO theory will be familiar to readers from almost every quantum chemistry textbook. Quantum chemical models can also be useful to analyze the details of chemical processes such as bond formations and reaction pathways. Solving the Schrödinger equation allows for detailed models of the interactions between electrons to be constructed, which are fundamental to chemistry.[2] This allows for the calculation of the electronic properties of molecules based on first principles. Spectroscopic observables, such as the energies of spectral peaks and their intensities, can also be predicted.

Hartree-Fock theory

The widely used Hartree-Fock method is the quantitative realization of molecular orbital theory. This method can be used to iteratively refine molecular orbitals to yield a self-consistent set of molecular orbitals and calculate the total energy of the system. The Hartree-Fock method is based upon three major approximations:

1. The Born-Oppenheimer approximation, which is the assumption that the motion of the electrons and nuclei can be separated. In practice this means fixing the positions of the nuclei, and calculating the motion of the electrons based on those nuclear coordinates.

2. The mean-field approximation, which is the assumption that the repulsion between electrons can be calculated by assuming that each electron moves within the average electrostatic field of all the other electrons in the system.
3. The single-reference approximation, where a single slater determinant is used to represent the state of the system. The Aufbau principle is used to determine orbital occupancies.

The Hartree-Fock procedure self-consistently refines a set of initial guess orbitals. At convergence, the total energy of the system is minimized under the above approximations. The energies of the orbitals are also generated by this procedure.

The electron correlation problem

The inaccuracy introduced by approximations (2) and (3) above is known as the electron correlation problem.[3] The electron correlation energy is defined as the difference between the exact energy of the chemical system and the Hartree-Fock energy, in a given basis set:[4,5]

$$E_{corr} = E_{exact} - E_{HF}$$

Although HF usually recovers more than 99% of the total energy of molecular systems, the remaining part of the energy is highly sensitive to the electrons' chemical environment, and electron correlation changes significantly during chemical processes.

A full configuration-interaction (FCI) calculation, in which all ways in which the electrons can occupy all orbitals are considered together in linear combination, gives the exact energy of the system,[6] as it is able to correct both approximations (2) and (3) above. The inaccuracy introduced by approximation (1) is usually negligible for most chemical processes. However, the computer time and memory required for a FCI calculation scale factorially, making FCI calculations infeasible for most chemically interesting systems. FCI is one of a number of 'multireference' methods, which relax constraint (3) above, and allow the wavefunction to be expressed as a linear combination of different orbital occupancies.[7,8]

It is possible to conceptually separate this correlation energy into 'dynamic' and 'static' components. Broadly speaking, the dynamic correlation energy corresponds to correcting for approximation (2) above – the mean-field approximation. Electrons move synchronously in order to maximize their distance from each other. The mean-field approximation employed in the derivation of the Hartree-Fock procedure assumes independent motion and that therefore systematically underestimates the distances between electrons, and thus overestimates electron repulsion energies. This correlated motion of the electrons is known as "dynamic correlation", and is highest between electrons that are paired within the same orbital but occurs between all electrons in a system (as all electrons interact). Long-range attraction between well separated neutral atoms or molecules is exclusively due to dynamic electron correlation, but may be referred to using terms such as van der Waals interactions, dispersion interactions, or instantaneous dipole–induced dipole interactions.

Non-dynamical correlation energy – also known as "static" correlation – arises when correcting for approximation (3) above. It occurs when there are near-degeneracies between the occupied and unoccupied orbitals of a chemical system. In such situations, a model that attempts to represent the wavefunction based upon a single electronic configuration breaks down. While a number of different ways of defining and calculating static correlation energies have been proposed,[9-15] it is most common to define it as the difference in energy between a wavefunction that combines all possible ways of arranging electrons within the valence orbitals and the Hartree-Fock wavefunction. However, since the number of ways of arranging electrons within the valence orbitals becomes exponentially large with respect to molecular size, such calculations are impractical for systems bigger than about 6 small atoms. For this reason it is helpful to use more approximate methods to recover static correlation energy.[15]

Post-Hartree-Fock methods

A general strategy to generate a pragmatic subset of the FCI orbital occupancies and thus achieve better scaling for large systems is to place a limitation on the amount of deviation from the HF orbital occupancies allowed. For example, creating a linear combination of orbital occupancies that are generated by moving one or two electrons from the original HF occupancies is known as configuration-interaction singles and doubles (CISD).[16] Triple and quadruple excitations can likewise be included also giving CISDT, CISDTQ etc. These methods rely on the fact that it is the configurations most similar to the HF result that tend to make the greatest contributions to the overall energy, and thus neglect of dissimilar configurations creates the least error in the energies. A more popular variant of the CISD type methods are the coupled cluster family of methods (CCSD, CCSDT, CCSDTQ, etc)[17,18] in which excitations are generated via an exponential excitation operator ansatz. Coupled-cluster methods usually give more accurate energies than their CI counterparts without a substantial increase in the runtime.[19,20]

The complete active space self-consistent field (CASSCF) method takes a different approach, and restricts – "freezes" – the occupation of many orbitals to that of the HF wavefunction, but allows a subset of the orbitals – the "active space" – full flexibility as to their electron occupations, allowing all possible electron occupations of those orbitals to be included.[21-24] The method also optimizes the shape of all orbitals in the system to give the lowest energy possible. This method scales factorially with regard to the number of orbitals included in the active space, and as a result 16 orbitals is usually about the practical limit for the size of the CASSCF active space. The selection of orbitals for inclusion in the active space is a somewhat arbitrary process, and many researchers simply include as many orbital near to the HOMO and LUMO as possible.

Perturbation theory provides an alternative approach to recovering 'missing' correlation energy. The most commonly used perturbation theory is second-order Møller–Plesset perturbation theory, known as MP2 when applied to a

Hartree-Fock reference.[25,26] When applied to a CASSCF reference, the method is known as complete active space second-order perturbation theory (CASPT2), or second-order multireference Møller–Plesset perturbation theory (MRMP2)[27-31] depending on the specific implementation. A different form of perturbation theory may be applied to account for higher excitations missing from truncated CI and CC wavefunction expansions, giving CISD(T) and CCSD(T).[20,32-42] Table 1 lists the most commonly used *ab initio* methods.

Table 1: Most commonly used *ab initio* computational methods. Going down the table increases runtime and memory usage, but also (usually) accuracy. N, represents the size of the 1-electron basis set used in the calculation, thus HF scales with the 4th power of the size of the basis set.

	Method	Orbitals used	Non-iterative energy correction	Scaling factor
Single reference	HF	HF	No	N^4
	MP2	HF	Yes	N^5
	CCSD	HF	No	N^6
	CCSD(T)	HF	Yes	N^7
Multi-reference	CASSCF	CASSCF	No	N^4 * *Factorial(active space)*
	MRMP2	CASSCF	Yes	N^5 * *Factorial(active space)*
	FCI	Any	No	*Factorial(N)*

An alternative approach is that provided by density function theory (DFT). Hohenberg and Kohn demonstrated that the ground state electronic energy is fully determined by the electron density.[43] However, the functional connecting these two values is not known, and many approximate functionals have been designed. One of the most popular functionals is B3LYP,[44] which mixes the HF energy and the energy of two other functionals.[45,46] During the DFT procedure the Kohn-Sham orbitals[47] are iteratively optimized based on the selected functional. Time-dependent density functional theory (TDDFT) is the application of DFT procedures to the excited states of a system.[48] A major attraction of DFT methods is that the scaling and runtime of the calculations are often similar to HF but usually provide more accurate energies.

Example system – the ground state of the beryllium atom

The beryllium atom will be used as a model system to illustrate the concepts and methods described above. According to the Aufbau principle, electrons are paired in both the *1s* and *2s* orbitals, giving $1s^2 2s^2$ as the ground state configuration, as a HF calculation confirms. However, a CASSCF calculation within an active space of the valence orbitals (i.e. the *2s* and *2p* orbitals) yields a state comprised of 4 electronic configurations. The $1s^2 2s^2$ configuration has a coefficient of 0.95 while the three $1s^2 2p_x^2$ configurations have coefficients of 0.18.

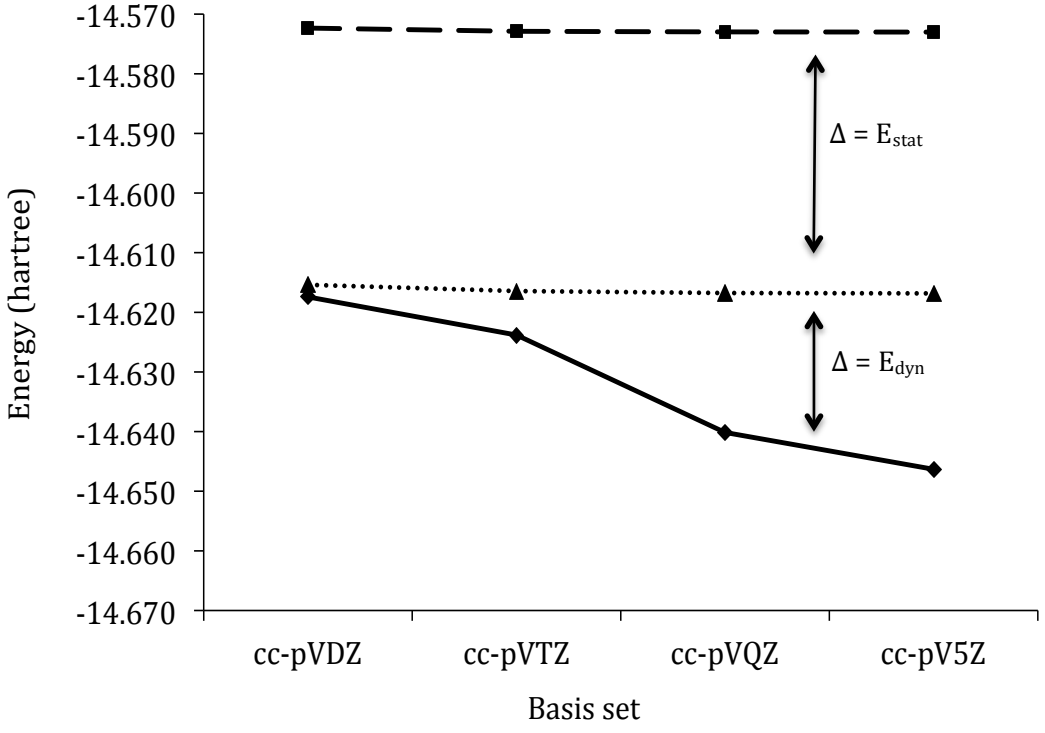


Figure 1: Energies of the ground state of Be in basis sets of increasing size. FCI - solid line; CASSCF (valence) - dotted line; HF - dashed line. Basis sets improve in quality from left to right, and are the correlation-consistent series developed by Dunning.[49]

Figure 1 shows the energy of the Be atom calculated using the FCI, CASSCF (with the valence orbitals as the active space), and HF methods. The HF and CASSCF lines are almost flat because the convergence of these methods toward the CBS limit is exponential with regard to the size of the one-electron basis,[50,51] while the FCI energy does not appear to have fully converged to the CBS limit even at cc-pV5Z and the rate of convergence of this method is polynomial with regard to

size of the one-electron basis.[52] The static correlation energy of 1.2 eV, given by the difference between the CASSCF and HF results, changes by less than 2% between the smallest and largest basis. We shall explore the relationship between static correlation energy and basis set in more detail in chapter 1.

Table 2: Energy of Be atom in the cc-pV5Z basis for various methods. CASSCF is in full valence region. The *1s* core orbital is not frozen in other calculations.

Method	Correlation energy recovered (eV)	Percentage of correlation energy recovered	Percentage of total energy recovered
HF	0	0%	99.50%
CASSCF	1.2	60%	99.80%
MP2	1.5	75%	99.87%
MRMP2	1.92	96%	99.98%
CCSD	1.98	99%	99.996%
CCSD(T)	1.99	99.95%	99.9998%
FCI	2.0	100%	100%

Table 2 compares the performance of different computational methods on this system. We see from this data that a Hartree-Fock calculation on Be yields 99.50% of the of the FCI energy, though that result has an error in the absolute energy of 2 eV (193 kJ mol^{-1}).

Comparing the CASSCF results to HF and FCI results shows that there is 1.2 eV of static correlation energy and 0.8 eV of dynamic correlation energy in this system. The MP2 calculation, which uses second-order perturbation theory to correct the energy of the Hartree-Fock starting point, recovers 75% of the total correlation energy neglected by HF. The MP2 method does not itself separate static and dynamic correlation. The MRMP2 calculation recovers 96% of the correlation energy and gives an absolute energy in error by 0.08 eV (7.7 kJ mol^{-1}). On this system the CCSD and CCSD(T) methods recover more than 99% of the correlation energy, and yield absolute energies accurate to within 0.02 eV (2 kJ mol^{-1}).

References

(1) Gil, V. *Orbitals in Chemistry*; Cambridge University Press: Cambridge, 2000.

(2) Levine, I. N. *Quantum Chemistry*; Fourth edition ed.; Prentice-Hall: New Jersey, 1991.

(3) Tew, D. P.; Klopper, W.; Helgaker, T. *J. Comput. Chem.* **2007**, *28*, 1307.

(4) Lowdin, P. O. *Adv Chem Phys* **1959**, *2*, 207.

(5) Pople, J. A.; Binkley, J. S. *Mol Phys* **1975**, *29*, 599.

(6) Szabo, A.; Ostlund, S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover: New York, 1989.

(7) Roos, B. O. In *Adv Chem Phys*; Lawley, K. P., Ed.; Wiley Interscience: New York, 1987; Vol. 69, p 339.

(8) Schmidt, M. W.; Gordon, M. S. *Annu Rev Phys Chem* **1998**, *49*, 233.

(9) Krylov, A. I.; Sherrill, C. D.; Byrd, E. F. C.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 10669.

(10) Krylov, A. I.; Slipchenko, L. V.; Levchenko, S. V. In *Electron Correlation Methodology*; American Chemical Society: 2007; Vol. 958, p 89.

(11) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2007**, *127*.

(12) Tsuchimochi, T.; Scuseria, G. E. *J. Chem. Phys.* **2009**, *131*.

(13) Chan, G. K. L.; Sharma, S. In *Solving the Schrodinger Equation: Has Everything Been Tried?*; Popelier, P., Ed.; Imperial College Press: London, 2011, p 43.

(14) Avella, A., Mancini, F. *Strongly Correlated Systems: Numerical Methods*; Springer, 2013; Vol. 176.

(15) Crittenden, D. L. *J. Phys. Chem. A* **2013**, *117*, 3852.

(16) Sherrill, C. D.; Schaefer, H. F. *Adv Quantum Chem* **1999**, *34*, 143.

(17) Shavitt, I.; Bartlett R. *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*; Cambridge University Press, 2009.

(18) Kummel, H. G. In *Recent progress in many-body theories: Proceedings of the 11th international conference.*; Bishop, R. F.; Brandes, T.; Gernoth, K. A.; Walet, N. R.; Xian, Y., Ed.; World Scientific Publishing: Singapore, 2002, p 334.

(19) Bartlett, R. J. In *Modern Electronic Structure Theory*; 2nd ed.; Yarkony, D.R., Ed.; World Scientific: 1995, p 1047.
(20) Hirata, S.; Nooijen, M.; Bartlett, R. J. *Chem. Phys. Lett.* **2000**, *326*, 255.
(21) Abrams, M. L.; Sherrill, C. D. *J. Phys. Chem. A* **2003**, *107*, 5611.
(22) Azizi, Z.; Roos, B. O.; Veryazov, V. *Phys Chem Chem Phys* **2006**, *8*, 2727.
(23) Olsen, J. *Int J Quantum Chem* **2011**, *111*, 3267.
(24) Pulay, P. *Int J Quantum Chem* **2011**, *111*, 3273.
(25) Møller, Chr; Plesset, M. S. *Phys Rev* **1934**, *46*, 618.
(26) Pople, J. A.; Seeger, R.; Krishnan, R. *Int J Quantum Chem* **1977**, 149.
(27) Hirao, K. *Chem. Phys. Lett.* **1992**, *196*, 397.
(28) Hirao, K. *Chem. Phys. Lett.* **1992**, *190*, 374.
(29) Hirao, K. *Int J Quantum Chem* **1992**, 517.
(30) Nakano, H. *J. Chem. Phys.* **1993**, *99*, 7983.
(31) Nakano, H. *Chem. Phys. Lett.* **1993**, *207*, 372.
(32) Bozkaya, U.; Schaefer, H. F. *J. Chem. Phys.* **2012**, *136*.
(33) Kowalski, K. *Chem. Phys. Lett.* **2005**, *411*, 306.
(34) Kowalski, K.; Piecuch, P. *J. Chem. Phys.* **2004**, *120*, 1715.
(35) Lee, Y. S.; Kucharski, S. A.; Bartlett, R. J. *J. Chem. Phys.* **1984**, *81*, 5906.
(36) Lutz, J. J. *Development and Applications of Coupled-Cluster Methods and Potential Energy Surface Extrapolation Schemes*, PhD Thesis, Michigan State University, 2011.
(37) Piecuch, P.; Wloch, M. *J. Chem. Phys.* **2005**, *123*.
(38) Sauer, S. P. A.; Schreiber, M.; Silva, M. R.; Thiel, W. *J Chem Theory Comput* **2009**, *5*, 555.
(39) Watts, J. D.; Bartlett, R. J. *J. Chem. Phys.* **1994**, *101*, 3073.
(40) Watts, J. D.; Bartlett, R. J. *Chem. Phys. Lett.* **1995**, *233*, 81.
(41) Watts, J. D.; Bartlett, R. J. *Chem. Phys. Lett.* **1996**, *258*, 581.
(42) Hirata, S.; Fan, P. D.; Auer, A. A.; Nooijen, M.; Piecuch, P. *J. Chem. Phys.* **2004**, *121*, 12197.
(43) Hohenberg, P.; Kohn, W. *Phys Rev B* **1964**, *136*, B864.
(44) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(45) Becke, A. D. *Phys Rev A* **1988**, *38*, 3098.
(46) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys Rev B* **1988**, *37*, 785.

(47) Kohn, W.; Sham, L. J. *Phys Rev* **1965**, *140*, 1133.

(48) Runge, E.; Gross, E. K. U. *Physical Review Letters* **1984**, *52*, 997.

(49) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(50) Petersson, G. A.; Malick, D. K.; Frisch, M. J.; Braunstein, M. *J. Chem. Phys.* **2005**, *123*.

(51) Karton, A.; Martin, J. M. L. *Theor Chem Acc* **2006**, *115*, 330.

(52) Kutzelnigg, W.; Morgan, J. D. *J. Chem. Phys.* **1992**, *96*, 4484.

2. Optimal Composition of Atomic Orbital Basis Sets for Recovering Static Correlation Energies

Abstract

Static correlation energies (E_{stat}) are calculated in a range of basis sets for a chemically diverse collection of atoms and molecules. The reliability of a basis set in capturing E_{stat} is assessed according to the following: mean and maximum absolute deviations from near-exact E_{stat} estimates, monotonic convergence to the complete basis set limit, and ability to capture E_{stat} accurately independent of changes in geometry, molecular size, and electronic configuration. Within the polarization and correlation-consistent basis set series, triple-ζ basis sets are the smallest that can reliably capture E_{stat}. The cc-pVTZ basis set performs particularly well, recovering E_{stat} to chemical accuracy for all atoms and molecules in our data set. A series of customized basis sets are constructed by stripping polarization functions from, and swapping polarization functions among, existing basis sets. Basis sets without polarization functions are incapable of accurately recovering E_{stat}. Basis sets with a near-complete set of s, p, and d functions can approach chemical accuracy in maximum absolute error. However, this may be achieved at lower computational cost by using a well balanced triple-ζ basis set including *f* functions, along with a smaller number of *s*, *p*, and *d* functions. Recommended basis sets for calculating E_{stat} with increasing accuracy at increasing computational cost are 6-311G(2d,2p), cc-pVTZ, and cc-pVQZ stripped of *g* functions.

Introduction

The correlation energy of a chemical system is defined[1] as the difference between its exact electronic energy and its numerically exact Hartree–Fock energy:

$$E_c = E^{exact} - E^{HF} \quad (1)$$

In practice, it is not generally possible to calculate E^{exact}, but highly accurate approximations of E^{exact} may be obtained via full configuration interaction

expansion (FCI) in a near-complete basis set (CBS).[2–4] It is well-known that the computational cost of full CI energy calculations scales factorially with basis size,[3] whereas the energy converges slowly to the CBS limit.[5] Therefore, thermochemically accurate FCI/ CBS calculations are only feasible for systems with up to three atoms from the first and second rows of the periodic table.[6–9]

Truncation of the CI expansion by excitation level where appropriate helps decrease computational cost, but it remains challenging to calculate chemically accurate total energies for molecules with more than five or six non-hydrogen atoms.[10]

The slow convergence of the FCI energy with respect to basis size may be traced back to the difficulties associated with capturing the interelectronic cusps in the true wave function in a basis of one-electron Slater determinants.[5] Therefore, it is useful to consider partitioning the total correlation energy into "static" and "dynamic" components:

$$E_{\mathrm{c}} = E_{\mathrm{stat}} + E_{\mathrm{dyn}} \tag{2}$$

where the static contribution arises from near-degeneracies between occupied and unoccupied orbitals, and the dynamic term accounts for changes in short-range interactions due to the presence of interelectronic cusps in the true wave function that are poorly approximated by Hartree–Fock theory. Wave functions which capture the static correlation energy (e.g., CASSCF)[11–13] are constructed quite differently to those designed primarily to efficiently recover the dynamic correlation energy, for example, R12-based schemes,[14,15] and variational Monte Carlo wave functions incorporating Jastrow factors.[16]

A number of different ways of defining and calculating E_{stat} have been proposed,[11–13,17–28] of which the most general and widely used is:[22–28]

$$E_{\mathrm{stat}} = E^{\mathrm{CASSCF(val)/CBS}} - E^{\mathrm{R(O)HF/CBS}} \tag{3}$$

The factorial computational scaling of CASSCF limits its applicability to systems of up to 14 valence electrons in 14 valence molecular orbitals, regardless of the quality of the underlying atomic orbital basis. This restriction can be overcome

by resorting to more approximate E_{stat} models, for example, valence orbital optimized coupled cluster doubles (VOO−CCD),[22] the spin-flip method,[23] or full CI calculations within valence molecular orbitals constructed from a minimal molecule polarized atomic orbital basis.[28] These approximate E_{stat} models are expected to converge toward the complete basis set limit at the same rate as those calculated according to eq 3.

Irrespective of the E_{stat} model, substantial computational savings may be realized by first evaluating E_{stat} in a moderately sized but sufficiently large one-electron basis and then recovering E_{dyn} using explicitly correlated methods, incorporating functions of the interelectronic distance to reconstruct the wave function cusps. The above insight has motivated the recent development of a number of promising multi reference explicitly correlated methods (MR-R12)[15] such as MRMP2-F12,[29] CASSCF[2]$_{\text{R12}}$,[30] CASPT2-F12,[31] MRCI-F12,[32] and geminal-CASSCF.[33]

In principle, it is also possible to recover E_{dyn} using density functional theory but, in practice,[18,24,34−36] it is hard to derive and/or parametrize such functionals to avoid "double-counting" (i.e., unintentionally recovering some or all of E_{stat} concurrently with E_{dyn}).

Initial applications suggest that MR-R12 methods are so successful in capturing E_{dyn} that the largest remaining source of error in the total energy lies in the incompleteness of the one-electron basis used to calculate E^{HF} and E_{stat}.[15]

It is well-established that Hartree−Fock energies converge more rapidly to the complete basis set limit than exact energies. For the pc-*n* basis series, the leading term in the basis set truncation error decays exponentially with the number of primitive s functions,[37,38]

$$E_n^{\text{HF}} = E_\infty^{\text{HF}} + Ae^{-B\sqrt{n_s}} \tag{4}$$

These empirical results are consistent with a recent theoretical analysis which demonstrates root-exponential convergence of the Hartree–Fock energy of atomic orbitals expanded in an even-tempered Gaussian basis:[39]

$$E_{n_s}^{\mathrm{HF}} = E_{\infty}^{\mathrm{HF}} + Ae^{-\pi\sqrt{3n_s}} \tag{5}$$

A similar relationship has been observed for the cc-pVnZ basis series,[40]

$$E_{n}^{\mathrm{HF}} = E_{\infty}^{\mathrm{HF}} + Ae^{-B\sqrt{n}} \tag{6}$$

where n is the cardinality of the basis set. For correlation-consistent basis sets, n gives the number of contracted Gaussian-type functions describing the valence orbitals, and the identity of the highest angular momentum functions present in the basis for second and third row atoms (n = L_{max} for atoms Li – Ar). The number of primitive Gaussians comprising each contracted valence orbital function also increases with n. Therefore, n is a general measure of basis set quality that encompasses the overall ability of the basis to recover the electron–nucleus cusps in the wave function, provide spatial flexibility in molecular orbitals, and capture atomic polarization effects within a molecular environment.

Despite this rapid convergence behavior, to reliably obtain Hartree–Fock total energies to chemical accuracy requires pentuple-ζ basis sets. Fortunately, a range of highly accurate numerical approximations are available to reduce the complexity of this task, such as resolution of the identity, resolution of the operator, pseudospectral, and Cholesky decomposition.[10] This makes large basis Hartree–Fock calculations feasible for most organic molecules of practical interest.

Basis set incompleteness errors in E^{HF} arise predominantly from the description of the chemically uninteresting nucleus– electron cusps in the wave function.[39] Relative energies (e.g., reaction enthalpies, atomization, and ionization energies) provide a much more chemically relevant metric of basis set convergence and typically only require double or triple-ζ basis sets to achieve chemical accuracy, as cusp errors cancel in these cases.

Having accounted or controlled for basis set incompleteness errors in E^{HF}, the largest potential source of remaining error in E^{exact} is basis set incompleteness error in E_{stat}. Much less is known about the basis set convergence behavior of E_{stat} than E^{HF}, although it has been empirically established that CASSCF energies also converge toward the CBS limit with the same root-exponential dependence on the number of primitive basis functions as Hartree–Fock.[41]

Intuition suggests that it should be possible to recover E_{stat} to chemical accuracy employing only small- to medium-sized atomic orbital basis sets. First, E_{stat} is much smaller in magnitude than E^{HF} or $E^{\mathrm{CASSCF(val)}}$ so the same order of relative error will correspond to a much smaller absolute error. Second, static correlation energies arise from near-degeneracies between occupied and unoccupied orbitals, which are expected to be only weakly sensitive to the completeness of the underlying atomic orbital basis.

The aims of this chapter, arising from the observations and hypotheses above, are the following: (1) determine the optimal size and composition of atomic orbital basis sets for recovering E_{stat}, that is, establishing the requirements for obtaining E_{stat} to chemical accuracy with minimal computational cost; (2) elucidate the convergence behavior of static correlation energies; and (3) establish thermochemically accurate static correlation energies for a chemically diverse range of atoms and molecules.

Table 1. Truncation Schemes for Polarization Consistent Basis Sets

Name	Description
pc-0[*sp*] = pc-0	unmodified pc-0 basis
pc-0[spd_2]	pc-0 basis, plus *d* functions from pc-2 basis
pc-1[*sp*]	minimal pc-1 basis, stripped of all polarization functions
pc-1[*spd*] = pc-1	unmodified pc-1 basis
pc-1[spd_2]	minimal pc-1 basis, plus *d* functions from pc-2 basis
pc-1[spd_3]	minimal pc-1 basis, plus *d* functions from pc-3 basis
pc-2[*sp*]	minimal pc-2 basis, stripped of all polarization functions
pc-2[spd_1]	minimal pc-2 basis, plus *d* functions from pc-1 basis
pc-2[*spd*]	pc-2 basis, stripped of *f* functions
pc-2[spd_3]	minimal pc-2 basis, plus *d* functions from pc-3 basis
pc-2[*spdf*] = pc-2	unmodified pc-2 basis
pc-3[*sp*]	minimal pc-3 basis, stripped of all polarization functions
pc-3[spd_2]	minimal pc-3 basis, plus *d* functions from pc-2 basis
pc-3[*spd*]	pc-3 basis, stripped of *f* and *g* functions
pc-3[*spdf*]	pc-3 basis, stripped of *g* functions
pc-3[*spdfg*] = pc-3	unmodified pc-3 basis
pc-4[*sp*]	minimal pc-4 basis, stripped of all polarization functions
pc-4[*spd*]	pc-4 basis, stripped of *f* and *g* and *h* functions
pc-4[*spdf*]	pc-4 basis, stripped of *g* and *h* functions
pc-4[*spdfg*] ≈ pc-4	pc-4 basis, stripped of *h* functions

Table 2. Truncation Schemes for Correlation-Consistent Basis Sets

Name	Description
cc-pVDZ[*sp*]	minimal cc-pVDZ basis, stripped of all polarization functions
cc-pVDZ[*spd*] = cc-pVDZ	unmodified cc-pV(D+*d*)Z basis
cc-pVTZ[*sp*]	minimal cc-pVTZ basis, stripped of all polarization functions
cc-pVTZ[*spd*]	cc-pV(T+*d*)Z basis, stripped of *f* functions
cc-pVTZ[*spdf*] = cc-pVTZ	unmodified cc-pV(T+*d*)Z basis
cc-pVQZ[*sp*]	minimal cc-pVQZ basis, stripped of all polarization functions
cc-pVQZ[*spd*]	cc-pV(Q+*d*)Z basis, stripped of *f* and *g* functions
cc-pVQZ[*spdf*]	cc-pV(Q+*d*)Z basis, stripped of *g* functions
cc-pVQZ[*spdfg*] = cc-pVQZ	unmodified cc-pV(Q+*d*)Z basis
cc-pV5Z[*sp*]	minimal cc-pV5Z basis, stripped of all polarization functions
cc-pV5Z[*spd*]	cc-pV(5+*d*)Z basis, stripped of *f* and *g* and *h* functions
cc-pV5Z[*spdf*]	cc-pV(5+*d*)Z basis, stripped of *g* and *h* functions
cc-pV5Z[*spdfg*] ≈ cc-pV5Z	cc-pV(5+*d*)Z basis, stripped of *h* functions

Methods

To investigate the sensitivity of E_{stat} to basis set composition, a range of truncated polarization-consistent basis sets[37,42,43] and correlation-consistent basis sets[44–47] were constructed as detailed in Tables 1 and 2. Modifications to polarization functions are described for Li–Ar. Equivalent modifications were

performed for H and He involving polarization functions one unit lower in angular momentum.

Hartree–Fock and full valence CASSCF calculations were carried out in the GAMESS suite of quantum chemical software[48] for all atoms and molecules in Pople's G1 data set,[49] plus all reactants, products, and transition states in the DBH reaction database,[50] and the notoriously multireference ozone molecule. This collection contains 82 systems with nonzero E_{stat}, and will be referred to hereafter as our data set. Geometries of all systems in our data set are available as Supplementary Information.

Static correlation energies were also calculated for a series of dissociating diatomics (H_2, N_2, and F_2) to further confirm the ability of each basis set to accurately recover E_{stat} during bonding breaking and formation.

For comparison, the same calculations were also done using a selection of Pople basis sets: STO-3G,[51,52] 6-31G,[53–55] 6-31G(d, p),[53–55] 6-31G(2d, 2p),[53–56] 6-311G,[57,58] 6-311G(d, p),[57,58] 6-311G(2d, 2p),[56–58] and 6-311G(2df, 2pd).[56–58]

Results and Discussion

Finite Basis Estimates of E_{stat}

Static correlation energies calculated in the near-complete cc-pV5Z basis set provide thermochemically accurate correlation energies against which the performance of smaller basis sets are assessed:

$$\Delta E_{\text{stat}}^{\text{Z}} = E_{\text{stat}}^{\text{cc-pV5Z}} - E_{\text{stat}}^{\text{Z}} \qquad (7)$$

where Z denotes the smaller basis set.

The ability of each basis set to recover E_{stat} is measured by its mean and maximum absolute values of ΔE_{stat} across the data set. The size of each basis set is quantified by the number of primitive basis functions describing the carbon atom ($n_{\text{prim}}^{\text{C}}$). The carbon atom has been chosen as a prototype to illustrate the

relative sizes of different basis sets, as it is the one of the most common constituents of the molecules within our data set and intermediate in atomic size. Mean and maximum absolute ΔE_{stat} and $n_{\text{prim}}^{\text{C}}$ for all non-customized basis sets are given in Table 3.

Table 3. Mean and Maximum Absolute Errors Across Our Data Set (mE_{h}), Plus Number of Primitive Functions Describing the Carbon Atom for Each Basis Set

Z	mean $\lvert\Delta E_{stat}^{Z}\rvert$	max $\lvert\Delta E_{stat}^{Z}\rvert$	n_{prim}^{C}
pc-0	5.8	34.9	16
pc-1	1.2	7.3	28
pc-2	0.6	2.9	51
pc-3	0.1	0.3	92
pc-4	0.04	0.2	129
cc-pVDZ	1.2	6.4	34
cc-pVTZ	0.3	1.3	50
cc-pVQZ	0.04	0.3	77
cc-pV5Z	0	0	107
STO-3G	11.4	92.2	15
6-31G	6.1	40.9	22
6-311G	6.5	41.7	26
6-31G(*d*,*p*)	1.3	7.5	27
6-311G(*d*,*p*)	1.2	10.3	31
6-31G(*2d*,*2p*)	1.0	5.0	32
6-311G(*2d*,*2p*)	0.6	4.1	36
6-311G(*2df*,*2pd*)	0.4	2.6	43

From Table 3, we observe that split-valence basis sets without polarization functions (pc-0, 6-31G, 6-311G) cannot accurately capture the static correlation energy, regardless of the number of *s* and *p* functions. These basis sets produce E_{stat} estimates with mean absolute deviations around 6 mE_{h} and worst-case errors up to 40 mE_{h}.

Table 4. Thermochemically Accurate cc-pV5Z E_{stat} Values and Basis Set Truncation Errors[a]

	E_{stat}	ΔE_{stat}				E_{stat}	ΔE_{stat}		
	cc-pV5Z	pc-2	cc-pVTZ	6-311G (*2df, 2pd*)		cc-pV5Z	pc-2	cc-pVTZ	6-311G (*2df, 2pd*)
Be	43.8	-1.3	-0.2	-0.1	Mg	31.4	-0.6	-0.1	-0.1
B	34.7	-0.9	-0.3	-0.2	Al	24.2	-0.4	-0.2	-0.1
C	19.3	-0.5	-0.2	-0.2	Si	13.4	-0.2	-0.1	0.0
H_2	18.5	0.0	-0.1	-0.1	Li_2[b]	8.8	0.5	0.0	0.0
LiH[b]	16.4	-0.2	-0.1	-0.1	Na_2[b]	10.6	0.0	0.1	0.0
LiF[b]	15.0	-0.3	-0.2	-0.2	NaCl[b]	8.4	0.0	-0.2	-0.5
BeH	27.7	-1.3	-0.3	-0.4	CH	43.1	-0.8	-0.2	-0.2
NH	26.1	-0.2	-0.1	0.0	OH	25.0	-0.2	-0.1	0.0
FH	24.4	-0.2	-0.1	0.1	ClH	17.4	-0.1	-0.1	0.0
CH_2 ($^3B_1$)	38.8	-0.4	-0.2	-0.1	SiH_2 ($^3B_1$)	32.9	-0.2	-0.2	-0.1
CH_2 ($^1A_1$)	62.1	-1.3	-0.4	-0.2	SiH_2 ($^1A_1$)	51.3	-2.8	-0.4	-0.5
CH_3	59.4	-0.6	-0.3	-0.2	SiH_3	42.7	-0.4	-0.3	-0.2
CH_4	83.2	-0.8	-0.5	-0.5	SiH_4	55.6	-0.7	-0.5	-0.4
NH_2	49.1	-0.5	-0.2	-0.1	PH_2	34.8	-0.2	-0.1	-0.1
NH_3	75.0	-0.7	-0.6	-0.6	PH_3	48.4	-0.4	-0.2	-0.1
OH_2	53.7	-0.4	-0.4	-0.2	SH_2	34.4	-0.2	-0.2	0.0
CH_3OH[c]	118.8	-1.0	-0.7	-0.5	CH_3SH	104.7	-0.8	-0.5	-0.3
CH_3Cl	87.2	-0.8	-0.5	-0.4	FCI	37.7	0.1	0.2	0.6
C_2H_6	151.2	-1.2	-0.7	-0.9	Si_2H_6	96.5	-1.0	-0.7	-0.4
O_2H_2	107.5	-0.6	-0.4	-0.1	ClO	42.6	0.3	0.0	0.2
N_2H_4	135.1	-1.0	-0.9	-0.8	HOCl	66.4	-0.2	-0.2	0.1
O_2	104.3	0.1	0.1	0.6	S_2	46.0	0.8	0.1	0.4
F_2	79.1	-0.1	0.0	0.2	Cl_2	23.3	0.2	0.1	0.3
CN	151.9	-0.5	-0.1	0.3	HCO	125.8	-0.4	-0.3	0.1
HCN	151.4	-1.3	-0.6	-0.7	H_2CO	147.1	-1.9	-1.3	-2.0
N_2	148.5	-0.4	-0.3	0.0	P_2	91.5	1.6	0.2	0.5
C_2H_4	144.1	-1.0	-0.4	0.2	Si_2	80.0	2.9	0.1	0.8
C_2H_2	148.0	-0.9	-0.3	0.3	SiO	122.1	-0.8	-0.2	0.3
CO	131.8	-0.7	-0.4	-0.1	CS	104.6	0.1	-0.2	-0.1
NO	120.6	-0.2	0.0	0.4	SO	61.2	1.8	1.2	2.6
CO_2	176.3	-0.2	-0.1	0.6	SO_2	130.4	1.3	0.8	2.1
HOH‡	40.6	-0.2	-0.2	-0.3	O_3	237.3	-0.1	0.3	1.1
$HOCH_4$‡[d]	124.1	-1.0	-0.9	-1.0	SH	18.7	-0.1	-0.1	0.0
CH_3CH_2	128.9	-1.0	-0.6	-0.4	SH_3‡	41.1	-0.1	-0.1	0.0
HC_2H_4‡	148.3	-0.9	-0.4	0.1	HClH‡	37.4	0.5	0.2	0.4
N_2O	214.9	-0.2	0.0	0.6	CH_3F	93.8	-0.8	-0.5	-0.3
HN_2O‡	224.1	-0.2	0.3	1.0	CH_3FCl‡	112.3	-0.4	0.0	0.4
HN_2	129.9	-0.3	0.0	0.4	OH^-	22.5	-0.2	0.0	0.2
HN_2‡	158.2	-0.3	-0.1	0.3	FCH_3Cl^-‡	77.1	-1.2	-0.2	0.5
HNC	140.8	-1.1	-0.3	0.0	$HOCH_3F^-$‡	119.7	-1.2	-0.4	0.0
HCN‡	141.9	-0.7	-0.3	0.0	ClCH3Cl‡	80.2	-0.7	-0.1	-0.4

[a] All energies reported in mE_h. [b] Calculated using (2,2) active space, with unoccupied alkali metal atomic orbitals and doubly occupied halide atomic orbitals. [c] Calculated using (10,10) active space, constraining oxygen atomic orbitals to remain doubly occupied. [d] Calculated using (11,11) active space, constraining oxygen atomic orbitals to remain doubly occupied. ‡ Transition state.

Including a single shell of polarization functions significantly improves basis set quality, reducing the mean absolute error to <1.6 mE_h = 1 kcal/mol, commonly accepted as "chemical accuracy", except in the case of pc-0, for which no polarization functions are defined. Even supplementing pc-0 with polarization functions from pc-2 only slightly improves E_{stat} estimates, reducing the mean and maximum absolute errors to ~5 and 20 mE_h, respectively. Therefore, it is primarily the inability of this basis to recover appropriately sized molecular orbitals that limits E_{stat} accuracy.

Double-ζ basis sets of similar composition—pc-1, cc-pVDZ, 6-31G(*d*, *p*)—yield acceptable mean absolute errors. However, maximum errors in individual cases can reach over 7 mE_h, which is clearly well outside chemical accuracy.

Triple zeta basis sets—pc-2, cc-pVTZ, 6-311G(2*df*, 2*pd*)—produce significantly more accurate E_{stat} values than their double-ζ counterparts, halving the mean and maximum errors. Notably, within this class of basis sets, cc-pVTZ outperforms both pc-2 and 6-311G(2*df*, 2*pd*). Indeed, cc-pVTZ is the smallest basis set that yields chemically accurate static correlation energies for all molecules in our data set, as listed in Table 4.

From this table, it also appears that the greatest flexibility in atomic orbital basis is required to describe oxygen in a range of molecular environments, with all triple-ζ basis sets struggling to recover E_{stat} to chemical accuracy for H_2CO, SO, and SO_2.

Table 5. Mean Absolute Errors, Maximum Errors (mE_h) in Static Correlation Energies and Primitive Basis Function Counts for the Carbon Atom for Truncated pc-*n* and cc-pV*n*Z Basis Sets

	mean $\lvert\Delta E_{stat}\rvert$					mean $\lvert\Delta E_{stat}\rvert$			
	[*sp*]	[spd]	[sdpf]	[spdfg]		[*sp*]	[spd]	[spdf]	[spdfg]
pc-1	6.8	1.2			cc-pVDZ	6.4	1.2		
pc-2	6.1	0.8	0.6		cc-pVTZ	6.2	0.5	0.3	
pc-3	6.2	0.3	0.05	0.05	cc-pVQZ	6.4	0.3	0.05	0.04
pc-4	6.1	0.2	0.04	0.04	cc-pV5Z	6.2	0.3	0.00	0
	max $\lvert\Delta E_{stat}\rvert$					max $\lvert\Delta E_{stat}\rvert$			
pc-1	38.8	7.3			cc-pVDZ	39.9	6.4		
pc-2	41.7	3.4	2.9		cc-pVTZ	41.4	2.7	1.3	
pc-3	40.5	1.7	0.3	0.3	cc-pVQZ	40.6	2.0	0.3	0.3
pc-4	39.7	1.6	0.2	0.2	cc-pV5Z	40.2	1.8	0.03	0
	n_{prim}					n_{prim}			
pc-1	23	28			cc-pVDZ	29	34		
pc-2	34	44	51		cc-pVTZ	33	43	50	
pc-3	49	69	83	92	cc-pVQZ	39	54	68	77
pc-4	60	90	111	129	cc-pV5Z	48	68	89	107

The pc-2 basis is generally less accurate than cc-pVTZ and produces particularly poor estimates of E_{stat} for the second-row containing molecules Si_2, P_2, and SiH_2 ($^1A_1$), in addition to the oxygen-containing molecules discussed above. This result is surprising, as the molecular data set used in constructing the pc-2 basis contained all three of these molecules.[42]

Quadruple and pentuple-ζ basis sets provide very accurate estimates of E_{stat} but at a correspondingly high computational cost. The rapid convergence of E_{stat} with respect to basis set size suggests that higher angular momentum functions do not contribute significantly to the accuracy of the results, and so they may be omitted to improve computational performance.

Optimal Basis Composition

To quantify the relationship between basis set composition, accuracy, and computational cost, mean and maximum absolute errors for a selection of truncated basis sets are reproduced in Table 5, along with primitive basis function counts.

As observed above, basis sets without polarization functions do not accurately capture the static correlation energy. Mean and maximum absolute errors for [*sp*] truncated polarization-consistent and correlation-consistent basis sets are very similar to those of pc-0, 6-31G and 6-311G. Therefore, the remainder of this discussion will focus on basis sets including at least one shell of polarization functions.

In the pc-*n* basis series, starting from pc-1, the greatest accuracy gains can be achieved for the least computational cost by first increasing the number of valence and first polarization shell functions, that is, moving down the [*spd*] basis series to pc-3[*spd*], before adding an additional shell of *f* functions to ensure sub-mE_h accuracy for all molecules and mean absolute errors less than 100 μE_h.

The correlation-consistent basis sets tell a slightly different story. Starting from cc-pVDZ, it is again best to initially increase the number of valence and first polarization shell functions to cc-pVTZ[*spd*]. From there, the mean absolute error data suggest that it is almost equally advantageous to add an additional layer of polarization functions (cc-pVTZ) or increase the number of valence and first polarization shell functions (ccpVQZ[*spd*]). However, adding *f* functions is more useful than increasing the number of *s*, *p*, and *d* functions in reducing the maximum absolute error. Physically, this means that atoms need the flexibility to adopt a range of different and highly asymmetric shapes to fit the large range of molecular environments encountered within our data set. It is more important to allow a larger range of shapes than further refine the size of the cc-pVTZ[*spd*] orbitals.

Analogous to the pc-*n* series, sub-mE_h accuracy for all molecules is guaranteed by the cc-pVQZ[*spdf*] basis. Additional valence and polarization functions beyond those contained within pc-3[*spdf*] and cc-pVQZ[*spdf*] do not improve the accuracy of calculated E_{stat} values enough to justify the increase in computational cost.

For each column in Table 5, excluding [*sp*], the number of valence orbital and polarization functions increase simultaneously as the quality of the basis set improves. To decouple these effects, we have constructed a series of basis sets with swapped polarization functions, whose performance is summarized in Table 6.

Table 6. Mean and Maximum Absolute Errors (mE_h) in Static Correlation Energies for Polarization Function Swapped Truncated pc-*n* Basis Sets

	mean $\lvert\Delta E_{stat}\rvert$		
	pc-1	pc-2	pc-3
[spd$_1$]	1.2	1.3	
[spd$_2$]	1.0	0.8	0.5
[spd$_3$]	0.9	0.5	0.3
	max $\lvert\Delta E_{stat}\rvert$		
[spd$_1$]	7.3	7.7	
[spd$_2$]	3.3	3.4	3.2
[spd$_3$]	3.0	2.3	1.7

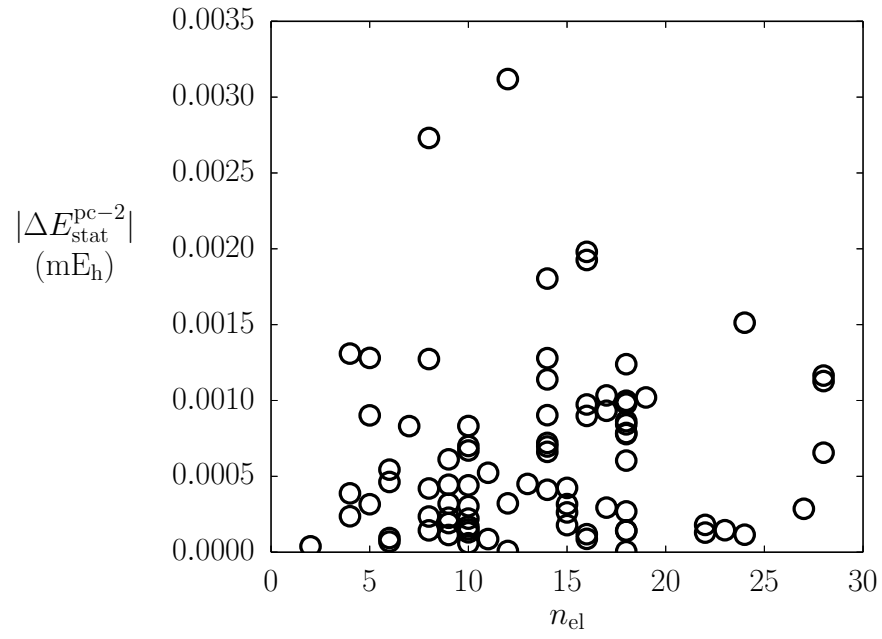
Basis sets which contain the same number of split valence functions and polarization functions, viz., pc-1[spd_2] and pc-2[spd_3], tend to provide the optimal balance between basis size and accuracy. From Table 6, it appears that improving the ability of basis sets to capture atomic polarization within the molecular environment plays a major role in increasing accuracy, particularly in terms of reducing maximum errors in E_{stat}.

Dependence of ΔE_{stat} on Molecule Size, Geometry, and Electronic Configuration.

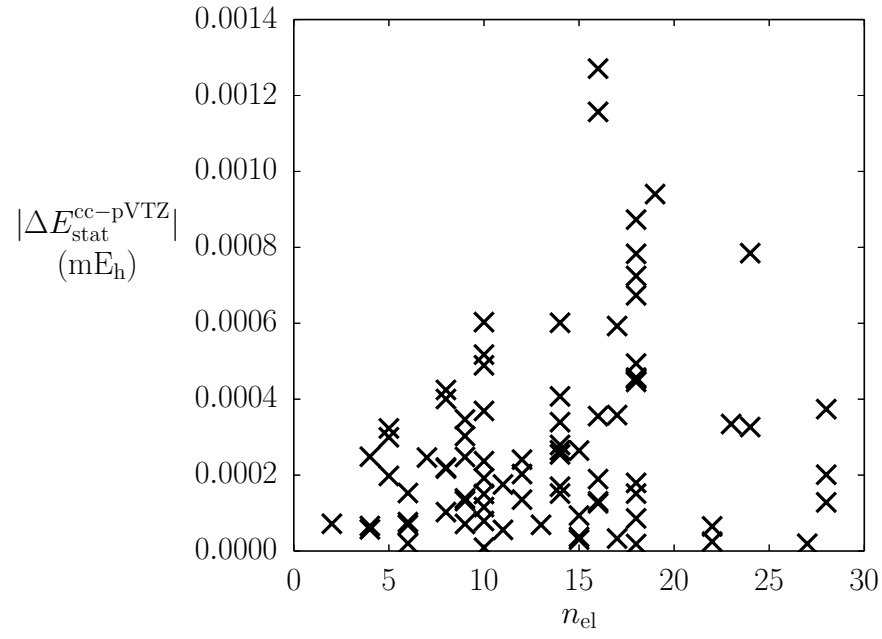
Mean and maximum absolute errors are a simple and robust metric for assessing the ability of each basis set to recover static correlation energies for a range of atomic and molecular systems. However, a more physically motivated criterion for acceptable performance is that ΔE_{stat} should be independent of molecule size, geometry, and bonding patterns. In other words, the atomic orbital basis should be flexible enough to describe each atom in a range of molecular environments.

The ability of the pc-2 basis to recover E_{stat} is largely independent of molecule size (measured as the number of valence electrons), as illustrated in Figure 1. Linear regression analysis indicates that there is no statistically significant

correlation between molecule size and error in E_{stat} values calculated in the pc-2 basis, with a near zero slope (9 ± 22 μE_{h}/ valence electron) and small R^2 value (<0.01). This is also true of the cc-pVTZ basis (slope = 10 ± 9 μE_{h} /valence electron and R^2 = 0.05). Reported uncertainties represent 95% confidence intervals.

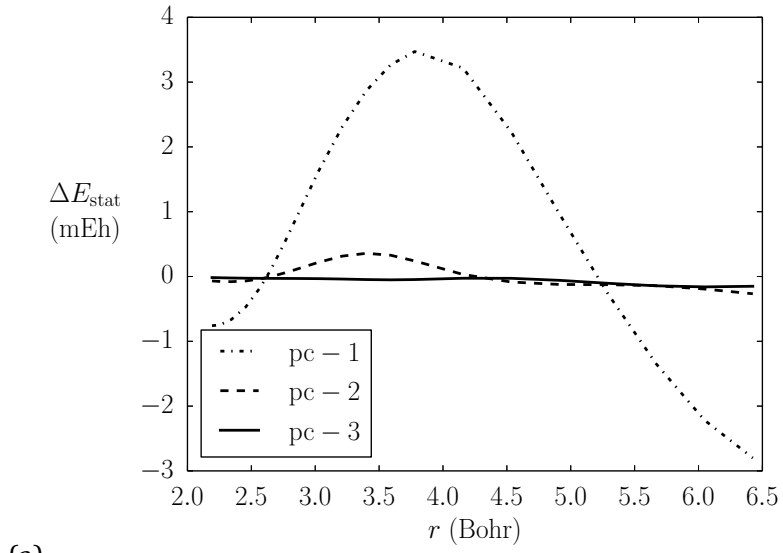


(a)

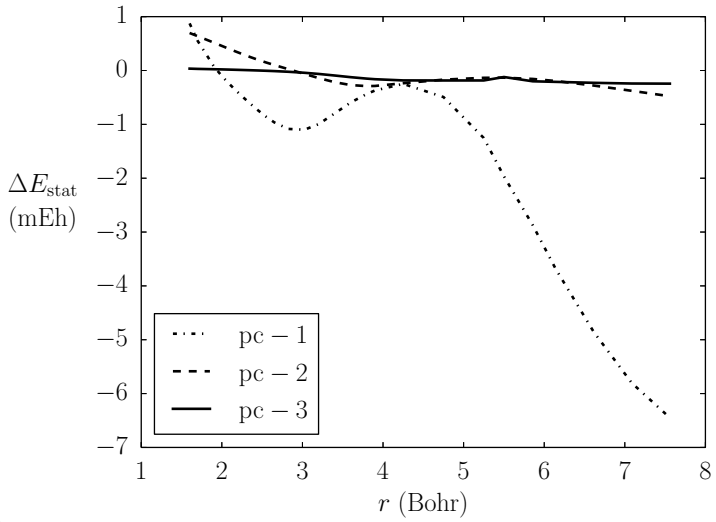


(b)

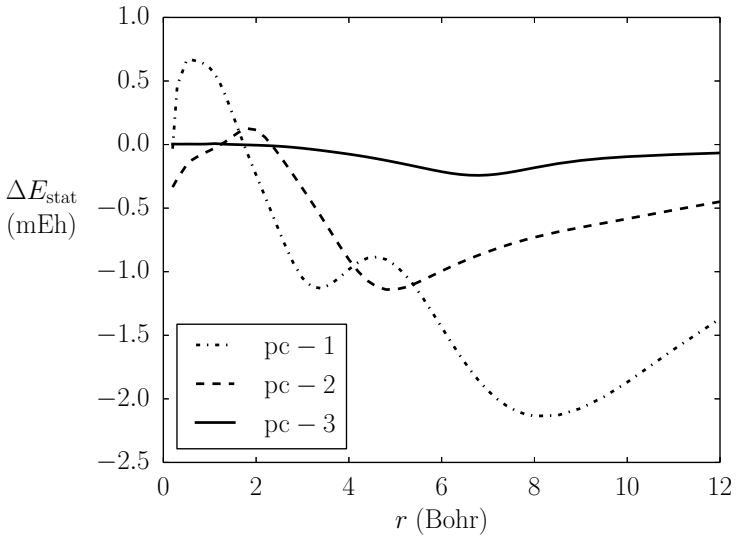
Figure 1. (a) $|\Delta E_{\mathrm{stat}}$ pc−2 | and (b) $|\Delta E_{\mathrm{stat}}$ cc−pVTZ| as a function of number of electrons for each atom, molecule, and transition state complex in our data set



(a)



(b)



(c)

Figure 2. Deviation of smaller basis static correlation energies from pc-4 reference values, ΔE_{stat}, as a function of bond length for a series of dissociating diatomics: (a) H_2, (b) N_2, (c) F_2.

In general, all basis sets of at least double-ζ quality show little to no correlation between ΔE_{stat} and molecular size. Hence, this is not as strict a criterion for acceptable performance as error threshold measures. However, it is reassuring that the errors calculated from our data set are independent of molecular size, so the conclusions about basis set quality drawn from our data set are expected to apply also to larger molecules. A complete set of linear regression parameters for all basis sets is provided in Appendix 1.

The ability of each basis set to capture E_{stat} in a range of molecular environments is most clearly illustrated by monitoring errors in E_{stat} as a function of bond length for a series of prototypical diatomics, as shown in Figure 2.

The pc-2 basis set provides the minimum flexibility required calculate E_{stat} to chemical accuracy across all bond lengths for H_2, N_2 and F_2.

On the basis of the dissociating diatomic curves in Figure 2, it is plausible to hypothesize that atomic orbital basis sets recover E_{stat} more effectively at their equilibrium geometries (~1.4 Bohr for H_2, ~2.1 Bohr for N_2, and ~2.7 Bohr for F_2). However, separate analysis of the G1 and DBH data sets reveals no significant difference in accuracy of E_{stat} estimates between equilibrium and transition state systems for all nonminimal basis sets. Therefore, triple-ζ basis sets provide an optimal compromise between accuracy and computational cost for molecules both at, and slightly displaced from, their equilibrium geometries.

The final physical test of basis set quality lies in the ability to capture E_{stat} accurately, independent of bonding patterns. As molecules with unsaturated bonding have larger E_{stat} than saturated systems, they may also have larger basis set incompleteness errors. Comparing saturated and unsaturated systems within our data set confirms that this is only true for basis sets smaller than triple-ζ. For triple-ζ and larger basis sets, ΔE_{stat} is independent of molecular electronic structure.

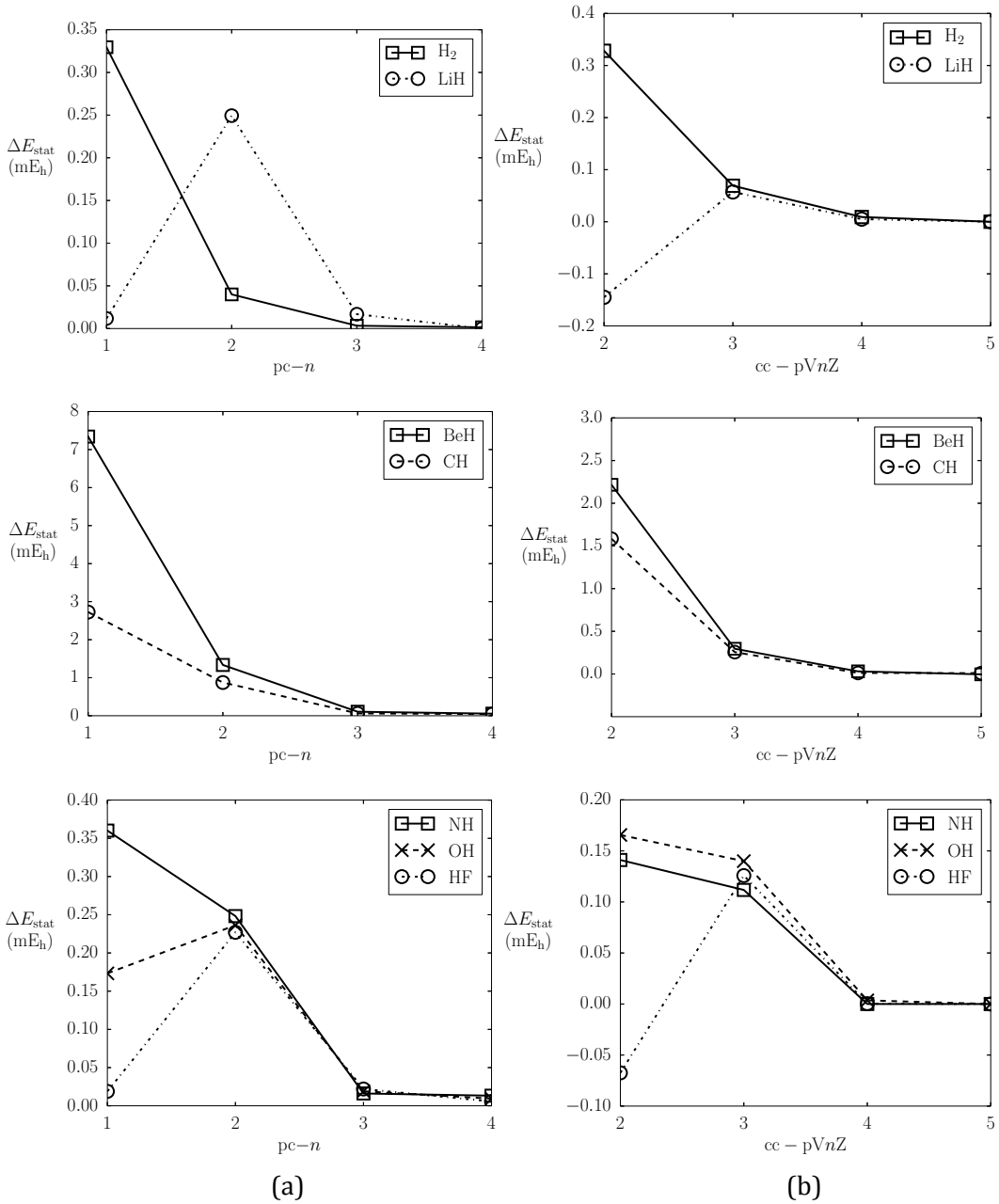


Figure 3. Deviation of static correlation energies from extrapolated complete basis set reference values for a series of hydrides, as a function of (a) polarization-consistent and (b) correlation-consistent basis size.

Basis Set Convergence Behavior of E_{stat}

The convergence of E_{stat} toward the complete basis set limit is illustrated for a series of diatomic hydrides in Figure 3. E_{stat} converges monotonically from pc-2 (Figure 3a) and cc-pVTZ (Figure 3b) to the complete basis set limit. This behavior is consistent across our complete data set (data in Appendix 1).

From smaller basis sets (pc-0, pc-1, cc-pVDZ), E_{stat} no longer converges monotonically to the CBS limit. In these cases, the CI expansion no longer primarily recovers the static correlation energy from qualitatively correct CASSCF orbitals, but it also compensates for basis set incompleteness errors. This complicated nonlinear coupling between CI and MO coefficients and the competing incentives to optimize both the orbitals and static correlation energy leads to the erratic convergence at the small basis end of each plot in Figure 3.

Similar basis set convergence patterns have also been observed in high level CCSD(T) calculations of spectroscopic parameters.[59] This suggests that the inadequacy of double-ζ basis sets leading to competing effects in determining CI/CC coefficients will show up clearly in any metric more sensitive than the total CASSCF energy. Further, it has long been observed that double-ζ basis sets are too small to be used reliably in basis set extrapolation schemes, even for total energies.[59]

Convergence of CASSCF Wave Functions and Energies

The basis set convergence of E_{stat} provides a useful diagnostic for monitoring convergence of individual CASSCF calculations within a basis set series. Convergence of all CASSCF calculations, not just full-valence, can be complicated by strong nonlinear coupling between configuration interaction (CI) and molecular orbital (MO) coefficients.[60] One surprising consequence of this is that convergence of the CASSCF energy with respect to basis set size for a given molecule and basis set series does not imply convergence of its configuration interaction and molecular orbital coefficients.

A less surprising consequence is that individual medium-to-large basis CASSCF calculations often converge to solutions that represent only local minima in CI and MO coefficient space.

The simplest solution to this problem is to use a smaller basis set. Indeed, CASSCF calculations in a minimal atomic basis always converge to the global

minimum, as molecular orbitals are completely defined by the preceding Hartree−Fock procedure, so there is no coupling between CI and MO coefficients. Unfortunately, minimal atomic orbital basis sets (e.g., STO-3G) provide unreliable estimates of E_{stat}, with mean and maximum absolute errors of ~12 and 93 mE_{h}, respectively.

However, the qualitatively correct valence molecular orbitals produced by minimal atomic orbital basis sets provide a useful starting point for finding global minimum CASSCF solutions in larger basis sets. To enable universal expansion of molecular orbitals in different atomic orbital bases, the supervisor constructed a small utility program[61] to perform least-squares fitting of one atomic orbital basis to another and subsequent expansion of the original molecular orbitals in terms of the new fitted basis functions.

This approach almost exactly reproduces molecular orbitals when expanding from a small AO basis to a larger AO basis. The BasisFit utility may also be reliably used to interconvert molecular orbitals between basis sets of the same quality but different basis set families, for example, pc-2 ↔ cc-pVTZ ↔ 6-311G(2*df*, 2*pd*). However, it is less reliable when fitting from a larger basis to a smaller one, because the smaller basis lacks the flexibility to exactly reproduce the functions in the larger AO basis, and contributions to the molecular orbitals from higher angular momentum functions than available in the smaller AO basis must be neglected. The resulting molecular orbitals are therefore slightly distorted and no longer orthogonal.

In practice, this makes it easier to converge CASSCF calculations by starting from smaller basis molecular orbitals, even if converged molecular orbitals from a larger basis are available. Therefore, to reliably obtain converged CASSCF molecular orbitals in medium to large basis sets, we advocate the following bootstrapping procedure:

- obtain molecular orbitals in minimal AO basis e.g. STO- 3G (note that CASSCF and Hartree−Fock orbitals are equivalent for most molecules whose orbitals are constrained by symmetry).

- expand molecular orbitals from minimal AO basis in a slightly larger AO basis (e.g., pc-0, as described above).
- optimize valence molecular orbitals during CASSCF calculation.
- expand these molecular orbitals in a yet larger AO basis (e.g., pc-0 → pc-1, pc-1 → pc-2, etc).
- repeat last two steps until E_{stat} is obtained with desired accuracy.

This procedure may also be useful to expedite the convergence of VOO–CCD calculations, which also involve concurrent orbital optimization and determination of coupled cluster coefficients.[22]

In this work, we have taken great care to ensure that CASSCF wave functions and energies are fully converged to their global minima. The CASSCF energies, in Appendix 1, may therefore be used for validating approximate E_{stat} models, as well as testing CASSCF convergence algorithms and strategies.

Accuracy of Near-CBS E_{stat} Values

The fully converged CASSCF energies in near-complete atomic orbital basis sets pc-4 and cc-pV5Z are expected to produce highly accurate static correlation energies, such that they may be considered exact for most practical purposes. However, it is important to verify this assertion by estimating the remaining basis set incompleteness error.

A range of strategies were employed to assess the accuracy of the pc-4 and cc-pV5Z E_{stat} values including:

- comparison of pc-4 and cc-pV5Z results
- comparison between quadruple and pentuple-ζ basis results
- comparison with extrapolated estimates of E_{stat} at the complete basis set limit

The basis set convergence behavior of E_{stat} across both the pc-(n-1) and cc-pVnZ basis set series is well-captured by a three-point extrapolation of the form:

$$E_{\text{stat},n} = E_{\text{stat},\infty} + Ae^{-B\sqrt{n}} \tag{9}$$

provided that all basis sets are of at least triple-ζ quality. RMS fitting errors are less than 21 mE_{h} across all molecules and basis sets in each basis set series. Here,

n represents the number of split-valence basis functions, and B is a fixed exponent optimized for each basis set series. $E_{\mathrm{stat},\infty}$ may also be denoted E_{stat} CBS.

In general, pc-*n* E_{stat} values converge faster to the CBS limit than cc-pV*n*Z, with optimal exponents of 3.2 and 2.5, respectively. This reflects the basis set convergence behavior observed for E^{HF} by Jensen.[37,42,43]

More sophisticated extrapolation schemes that depend on the number of primitive s-type basis functions are available for the pc-*n* basis series, but not for cc-pV*n*Z. Further, E_{stat}, as a relative energy, appears to exhibit basis set convergence behavior that is less sensitive to choice of extrapolation variable than its parent E^{HF} and $E^{\mathrm{CASSCF(val)}}$ absolute energies.

All quantities relevant to assessing the accuracy of pentuple-ζ results are presented in Table 7. Differences between various estimates of E_{stat} are quantified as:

$$\Delta E_{\mathrm{stat}}^{\mathrm{X,Y}} = E_{\mathrm{stat}}^{\mathrm{X}} - E_{\mathrm{stat}}^{\mathrm{Y}} \qquad (10)$$

with X and Y chosen as shown in Table 7.

Table 7. Mean and Maximum Absolute Deviations in E_{stat} (μE_{h}), between Pentuple-ζ Correlation Consistent and Polarization Consistent Basis Sets, between Quadruple-ζ and Pentuple-ζ Basis Sets within Each Series, and between Pentuple-ζ Basis Sets and Estimates of E_{stat} at the Complete Basis Set Limit Extrapolated from the Correlation Consistent Basis Set Series (cc-CBS) and the Polarization Consistent Basis Set Series (pc-CBS)

X	Y	mean $\lvert\Delta E_{stat}^{\mathrm{X,Y}}\rvert$		maximum $\lvert\Delta E_{stat}^{\mathrm{X,Y}}\rvert$	
		first row molecules	second row molecules	first row molecules	second row molecules
cc-pV5Z	pc-4	14	79	84	228
cc-pV5Z	cc-pVQZ	40	43	229	309
pc-4	pc-3	24	46	177	247
cc-CBS	cc-pV5Z	14	17	75	103
pc-CBS	pc-4	9	12	74	57

Differences between quadruple and pentuple-ζ basis results should provide an upper limit on the remaining variation between the pentuple-ζ results and the

complete basis set limit, but comparison of pc-4 and cc-pV5Z results reveals that these upper limits are exceeded for molecules containing second row atoms (Na–Ar). This implies that either the pc-*n* or cc-pV*n*Z basis set series for second row atoms possesses systematic deficiencies which prevent convergence to the true CBS limit.

This supposition can only be tested indirectly by comparing finite basis and numerically exact Hartree–Fock energies for a selection of atoms and diatomic molecules from our data set, per the results presented in Table 8. In all cases, the numerical HF energies were calculated to sub-μE_h accuracy. These values, along with the number of grid points and radial parameters used to generate them, are given in Appendix 1.

Table 8. Deviation of pc-4 and cc-pV5Z Hartree–Fock Energies (mE_h) from Their Numerically Exact Values

	ΔE^{HF}			ΔE^{HF}	
	pc-4	cc-pV5Z		pc-4	cc-pV5Z
H	0.000	0.005	He	0.005	0.055
Li	0.083	0.004	Na	8.311	0.236
Be	0.132	0.011	Mg	7.207	0.031
B	0.012	0.029	Al	6.386	0.081
C	0.039	0.051	Si	5.935	0.105
N	0.036	0.082	P	5.650	0.104
O	0.057	0.166	S	5.475	0.171
F	0.061	0.241	Cl	5.323	0.193
F-	0.213	4.028	Cl^-	5.468	2.020
Ne	0.055	0.328	Ar	5.202	0.171
H_2	0.002	0.022	Na_2	16.605	0.578
HF	0.058	0.374	NaCl	13.854	0.961
LiF	0.167	0.333	FCl	5.348	0.737
F_2	0.155	0.706	Cl_2	10.619	0.603
N_2	0.099	0.447	P_2	11.411	0.528
CO	0.108	0.313	CS	5.631	0.483

Overall, cc-pV5Z accurately recovers the exact HF energies, except for anions, where the cc-pV5Z energies are in error by up to 4 mE_h. These results reinforce the well-known conclusion that diffuse functions are required when modeling anions using correlation-consistent basis sets.

On the other hand, the pc-4 basis performs equally well for anions and uncharged species. Providing this additional flexibility in the valence region comes at the expense of accurately modeling the chemically uninteresting core region for second row atoms. The pc-4 Hartree–Fock energies for second-row atoms differ from the exact values by up to 8 mE_h. This error is additive as second row atoms are incorporated into molecules, but is expected to cancel when calculating relative energies (e.g., reaction enthalpies, activation energies, static correlation energies).

Indeed, the mE_h basis set incompleteness errors in E^{HF} for second-row atoms cancel to become only μE_h errors in E_{stat}. Nonetheless, as pentuple-ζ basis sets are capable of recovering E_{stat} to μE_h accuracy, this represents a significant and systematic deviation of the pc-4 E_{stat} values from the true CBS limit.

CBS-extrapolated static correlation energies provide a more realistic estimate of differences between pentuple-ζ results and the complete basis set limit. These results suggest that the pc-4 and cc-pV5Z results lie, on average, within 20 μE_h of the complete basis set limit, although this does not account for the systematic deficiencies in the polarization-consistent basis sets for second-row atoms. Basis set extrapolation predicts a maximum deviation from the CBS limit of 0.1 mE_h, whereas more conservative approaches indicate that the maximum error could be up to 0.3 mE_h.

Conclusions

Split valence basis sets without polarization functions cannot accurately capture the static correlation energy, regardless of the number of *s* and *p* functions. On the other hand, functions of high angular momentum—*g* and above—can be excluded from the basis set without appreciable loss of accuracy in E_{stat}. The sweet spot for the inclusion of polarization functions lies in the *d* and *f* functions. Chemical accuracy for all molecules may be achieved using a near complete basis of *s*, *p*, and *d* functions. However, this may be achieved at lower computational

cost by using a well-balanced basis set including *f* functions, along with a smaller number of *s*, *p*, and *d* functions.

For this reason, triple-ζ basis sets exhibit a near-optimal balance between accuracy and computational cost in recovering static correlation energies. In particular, the cc-pVTZ basis provides chemically accurate E_{stat} values for all molecules in our data set. The 6-311G(2*d*,2*p*) basis gives remarkably accurate results for its relatively small size but cannot guarantee chemical accuracy for all molecules.

Truncated quadruple-ζ basis sets stripped of their *g* functions provide the most cost-effective way of recovering E_{stat} to sub-mE_{h} accuracy and may be considered close enough to the complete basis set limit for most practical purposes.

References

(1) Löwdin, P. O. The Correlation Problem in Many-Electron Quantum Mechanics. 1. Review of Different Approaches and Discussion of Some Current Ideas. *Adv. Chem. Phys.* **1959**, *2*, 207– 322.

(2) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; McGraw-Hill: New York, 1989.

(3) Sherrill, C. D.; Schaefer H. F. In *The Configuration Interaction Method: Advances in Highly Correlated Approaches*; Per-Olov Löwdin, M. C. Z., John R. Sabin, Brandas, E., Eds.; Advances in Quantum¨ Chemistry; Academic Press, 1999; Vol. *34*; pp 143 – 269.

(4) O'Neill, D. P.; Gill, P. M. W. Benchmark Correlation Energies for Small Molecules. *Mol. Phys.* **2005**, *103*, 763–766.

(5) Kutzelnigg, W.; Morgan, J. D. Rates of Convergence of the Partial Wave Expansions of Atomic Correlation Energies. *J. Chem. Phys*. **1992**, *96*, 4484–4508.

(6) Feller, D. The Use of Systematic Sequences of Wave Functions for Estimating the Complete Basis Set, Full Configuration Interaction Limit in Water. *J. Chem. Phys.* **1993**, *98*, 7059–7071.

(7) Csaszar, A. G.; Tarczay, G.; Leininger, M. L.; Polyansky, O. L.; Tennyson, J.; Allen, W. D. *Spectroscopy from Space*; NATO Science Series; Springer, 2001; Vol. *20*; Chapter 19. Dream or Reality: Complete Basis Set Full Configuration Interaction Potential Energy Hypersurfaces, pp 317−339.

(8) Dutta, A.; Sherrill, C. D. Full Configuration Interaction Potential Energy Curves for Breaking Bonds to Hydrogen: An Assessment of Single-Reference Correlation Methods. *J. Chem. Phys.* **2003**, *118*, 1610−1619.

(9) Varandas, A. J. Extrapolation to the Complete Basis Set Limit without Counterpoise. The Pair Potential of Helium Revisited. *J. Phys. Chem. A* **2010**, *114*, 8505−8516.

(10) Sherrill, C. D. Frontiers in Electronic Structure Theory. *J. Chem. Phys*. **2010**, *132*, 110902.

(11) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M.; Elbert, S. T. Are Atoms Intrinsic to Molecular Electronic Wavefunctions? I. The FORS Model. *Chem. Phys.* **1982**, *71*, 41−49.

(12) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M. Are Atoms Intrinsic to Molecular Electronic Wavefunctions? II. Analysis of FORS Orbitals. *Chem. Phys.* **1982**, *71*, 51−64.

(13) Ruedenberg, K.; Schmidt, M. W.; Gilbert, M. M.; Elbert, S. T. Are Atoms Intrinsic to Molecular Electronic Wavefunctions? III. Analysis of FORS Configurations. *Chem. Phys.* **1982**, *71*, 65−78.

(14) Klopper, W.; Manby, F. R.; Ten-No, S.; Valeev, E. F. R12 Methods in Explicitly Correlated Molecular Electronic Structure Theory. *Int. Rev. Phys. Chem.* **2006**, *25*, 427−468.

(15) Kong, L.; Bischoff, F. A.; Valeev, E. F. Explicitly Correlated R12/F12 Methods for Electronic Structure. *Chem. Rev.* **2012**, *112*, 75− 107.

(16) Braida, B.; Toulouse, J.; Caffarel, M.; Umrigar, C. J. Quantum Monte Carlo with Jastrow-Valence-Bond Wave Functions. *J. Chem. Phys*. **2011**, *134*, 084108.

(17) Sinanoglu, O. Many-Electron Theory of Atoms, Molecules and their Interactions. *Adv. Chem. Phys*. **1964**, *6*, 315−412.

(18) Mok, D. K. W.; Neumann, R.; Handy, N. C. Dynamical and Nondynamical Correlation. *J. Chem. Phys*. **1996**, *100*, 6225−6230.

(19) Lee, M. S.; Head-Gordon, M. Extracting Polarized Atomic Orbitals from Molecular Orbital Calculations. *Int. J. Quantum Chem.* **2000**, *76*, 169–184.

(20) Lu, W. C.; Wang, C. Z.; Schmidt, M. W.; Bytautas, L.; Ho, K. M.; Ruedenberg, K. Molecule Intrinsic Minimal Basis Sets. I. Exact Resolution of Ab Initio Optimized Molecular Orbitals in Terms of Deformed Atomic Minimal-Basis Orbitals. *J. Chem. Phys.* **2004**, *120*, 2629–2638.

(21) Abrams, M. L.; Sherrill, C. D. Natural Orbitals as Substitutes for Optimized Orbitals in Complete Active Space Wavefunctions. *Chem. Phys. Lett.* **2004**, *395*, 227–232.

(22) Krylov, A. I.; Sherrill, C. D.; Byrd, E. F. C.; Head-Gordon, M. Size-consistent Wave Functions for Nondynamical Correlation Energy: The Valence Active Space Optimized Orbital Coupled-Cluster Doubles Model. *J. Chem. Phys.* **1998**, *109*, 10669–10678.

(23) Krylov, A. I.; Slipchenko, L. V.; Levchenko, S. V. In *Electron Correlation Methodology*; Wilson, A., Ed.; ACS Symposium Series; American Chemical Society: Washington, DC, 2007; Chapter 6. Breaking the Curse of the Non-Dynamical Correlation Problem: The Spin-Flip Method.

(24) Becke, A. D.; Johnson, E. R. A Unified Density-Functional Treatment of Dynamical, Nondynamical and Dispersion Correlations. *J. Chem. Phys.* **2007**, *127*, 124108.

(25) Tsuchimochi, T.; Scuseria, G. E. Strong Correlations via Constrained-Pairing Mean-Field Theory. *J. Chem. Phys.* **2009**, *131*, 121102.

(26) Chan, G. K. L.; Sharma, S. In *Solving the Schrodinger Equation: Has Everything Been Tried?*; Popelier, P., Ed.; Imperial College Press: London, 2011; Chapter 3. Solving Problems With Strong Correlation Using The Density Matrix Renormalization Group (DMRG), pp 43– 60.

(27) Avella, A., Mancini, F., Eds. *Strongly Correlated Systems: Numerical Methods*, 1st ed.; Springer Series in Solid-State Sciences; Springer, 2013; Vol. *176*.

(28) Crittenden, D. L. A Hierarchy of Static Correlation Models. *J. Phys. Chem. A* **2013**, *117*, 3852–3860.

(29) Ten-No, S. A Simple F12 Geminal Correlation in MultiReference Perturbation Theory. *Chem. Phys. Lett.* **2007**, *447*, 175–179.

(30) Torheyden, M.; Valeev, E. F. Universal Perturbative Explicitly Correlated Basis Set Incompleteness Correction. *J. Chem. Phys.* **2008**, *131*, 171103.

(31) Shiozaki, T.; Werner, H.-J. Second-Order Multireference Perturbation Theory with Explicit Correlation CASPT2-F12. *J. Chem. Phys.* **2010**, *133*, 141103.

(32) Shiozaki, T.; Knizia, G.; Werner, H.-J. Explicitly Correlated Multireference Configuration Interaction: MRCI-F12. *J. Chem. Phys.* **2011**, *134*, 034113.

(33) Varganov, S.; Martinez, T. J. Variational Geminal-Augmented Multireference Self-Consistent Field Theory: Two-Electron Systems. *J. Chem. Phys.* **2010**, *132*, 054103.

(34) Grimme, S.; Waletzke, M. A Combination of Kohn-Sham Density Functional Theory and Multi-Reference Configuration Interaction Methods. *J. Chem. Phys.* **1999**, *111*, 5645–5656.

(35) Leininger, T.; Stoll, H.; Werner, H.-J.; Savin, A. Combining Long-Range Configuration Interaction with Short-Range Density Functionals. *Chem. Phys. Lett.* **1997**, *275*, 151–160.

(36) Grafenstein, J.; Cremer, D. The Combination of Density Functional Theory with Multi-Configuration Methods - CAS-DFT. *Chem. Phys. Lett*. **2000**, 316.

(37) Jensen, F. Polarization Consistent Basis Sets: Principles. *J. Chem. Phys*. **2001**, *115*, 9113–9125.

(38) Jensen, F. Estimating the Hartree-Fock Limit from Finite Basis Set Calculations. *Theor. Chem. Acc.* **2005**, *113*, 267–273.

(39) McKemmish, L. K.; Gill, P. M. W. Gaussian Expansions of Orbitals. *J. Chem. Theory Comput.* **2012**, *8*, 4891–4898.

(40) Karton, A.; Martin, J. M. L. Comment on: ″Estimating the Hartree-Fock Limit from Finite Basis Set Calculations″ [Jensen F (2005) Theor Chem Acc 113:267]. *Theor. Chem. Acc*. **2006**, *115*, 330– 333.

(41) Petersson, G. A.; Malick, D. K.; Frisch, M. J.; Braunstein, M. The Convergence of Complete Active Space Self-Consistent-Field Energies to the Complete Basis Set Limit. *J. Chem. Phys.* **2005**, *123*, 074111.

(42) Jensen, F.; Helgaker, T. Polarization Consistent Basis Sets. V. The Elements Si–Cl. *J. Chem. Phys.* **2004**, *121*, 3463–3470.

(43) Jensen, F. Polarization Consistent Basis Sets. VI. The Elements He, Li, Be, B, Ne, Na, Mg, Al, Ar. *J. Phys. Chem. A* **2007**, *111*, 11198– 11204.

(44) Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *80*, 1007–1023.

(45) Woon, D. E.; Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum Through Argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(46) Woon, D. E.; Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. IV. Calculation of Static Electrical Response Properties. *J. Chem. Phys.* **1994**, *100*, 2975–2988.

(47) Dunning, T. H., Jr.; Peterson, K. A.; Wilson, A. K. Gaussian Basis Sets for Use in Correlated Molecular Calculations. X. The Atoms Aluminum Through Argon Revisited. *J. Chem. Phys*. **2001**, *114*, 9244– 9253.

(48) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; et al. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1363–1363.

(49) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, A. Gaussian-1 Theory: A General Procedure for Prediction of Molecular Energies. *J. Chem. Phys.* **1989**, *90*, 5622–5630.

(50) Zheng, J.; Zhao, Y.; Truhlar, D. G. Representative Benchmark Suites for Barrier Heights of Diverse Reaction Types and Assessment of Electronic Structure Methods for Thermochemical Kinetics. *J. Chem. Theory Comput.* **2007**, *3*, 569–582.

(51) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *J. Chem. Phys.* **1969**, *51*, 2657–2664.

(52) Hehre, W. J.; Ditchfield, R.; Steward, R. F.; Pople, J. A. Self-Consistent Molecular Orbital Methods. IV. Use of Gaussian Expansions of Slater-Type Orbitals. Extension to Second-Row Molecules. *J. Chem. Phys.* **1970**, *52*, 2769–2773.

(53) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in

Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(54) Dill, J. D.; Pople, J. A. Self-Consistent Molecular Orbitals Methods. XV. Extended Gaussian-Type Basis Sets for Lithium, Beryllium and Boron. *J. Chem. Phys*. **1975**, *62*, 2921–2923.

(55) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XXIII. A Polarization-Type Basis Set for Second-Row Elements*. J. Chem. Phys.* **1982**, *77*, 3654–3665.

(56) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-Consistent Molecular Orbital Methods 25. Supplementary Functions for Gaussian Basis Sets. *J. Chem. Phys.* **1984**, *80*, 3265–3269.

(57) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XX. A Basis Set for Correlated Wave Functions. *J. Chem. Phys.* **1980**, *72*, 650–654.

(58) McLean, A. D.; Chandler, G. S. Contracted Gaussian Basis Sets for Molecular Calculations. I. Second Row Atoms, Z=11–18. *J. Chem. Phys*. **1980**, *72*, 5639–5648.

(59) Temelso, B.; Valeev, E. F.; Sherrill, C. D. A Comparison of One-Particle Basis Set Completeness, Higher-Order Electron Correlation, Relativistic Effects, and Adiabatic Corrections for Spectroscopic Constants of BH, CH+, and NH. *J. Phys. Chem. A* **2004**, *108*, 3068–3075.

(60) Sherrill, C. D.; Krylov, A. I.; Byrd, E. F. C.; Head-Gordon, M. Energies and Analytic Gradients for Coupled-Cluster Doubles Model Using Variational Brueckner Orbitals: Application to Symmetry Breaking in O4+. *J. Chem. Phys.* **1998**, *109*, 4171–4181.

(61) BasisFit. http://sourceforge.net/projects/basisfit/

3. Cyclopropenium Cations Break the Rules of Attraction to Form Closely Bound Dimers

Abstract

The crystal structures of tris(ethylmethylamino)-cyclopropenium chloride and tris(diethylamino)-cyclopropenium iodide reveal the presence of closely-bound dicationic dimers formed from two closed-shell monomer units. The distances between the C_3 centroids of the staggered monomers are at the short end of those normally found in pi-stacked neutral arenes, let alone charged aromatic rings. Computational analysis reveals that short-range interactions are dominated by strong dispersion forces, enabling metastable dicationic dimers to form without covalent intermolecular bonding. Surrounding counterions then provide a background source of charge balance, imparting strong thermodynamic stability to the system. Additionally, these counterions form a weak but attractive electrostatic bridge between the monomer units, contributing to the surprisingly short observed intermolecular C_3–C_3 centroid distance.

Introduction

Ionic liquids (ILs) are chemically interesting materials that are finding increasing practical applications due to their useful physical properties such as neglible volatility, good conductivity, and an ability to dissolve a wide range of solutes.[1] These properties arise from the unusual chemical environment created by a mixture of ions in the liquid state; although ILs often consist of weakly-polar molecules, they usually behave like polar solvents. However, they can also exhibit non-homogeneous behaviour, with both hydrophilic (ion-rich) and hydrophobic (alkyl-rich) regions in a dynamic equilibrium.[2–6] Understanding the balance of intermolecular interactions within ILs that determine ion mobility, ion pairing and ion clustering, and their effects on the properties of ILs, is a fundamental problem of continuing interest.

We recently described the use of triaminocyclopropenium (TAC) cations, $[C_3(NR_2)_3]^+$, in ionic liquids.[7] These cations are geometrically and electronically

quite different from other aromatic cations commonly used for ionic liquids, such as imidazolium and pyridinium, exhibiting higher symmetry with a 3-membered all-carbon aromatic ring, and electron-donating amino substituents rather than N atoms incorporated into the aromatic ring. A high-lying HOMO also results in an unusually low and reversible oxidation potential.[8,9] Consequently, TAC salts have attracted much interest[10,11] since they were first reported in 1971 by Yoshida and Tawara.[12]

TAC cations also demonstrate unusually relatively weak short-range electrostatic interactions with surrounding counterions. This is manifest in the observation that halide counterions eschew interactions with the TAC cation, preferentially coordinating to other compounds instead. For example, Weiss has commented that "[t]ypically enough, the halides cannot be obtained in solvent-free form".[13] Chloride hydrates,[14,15] iodide-iodoacetylene[16] and iodide-iodoarene[17] adducts have all been isolated from tris(dialkylamino)-cyclopropenium (TDAC) halide salts. In particular, the dichloride hexahydrate structure shows essentially no distortion from its calculated gas phase structure, in contrast to those found with other cations.[14,15]

This apparent preference of TAC cations to coordinate to species other than counterions may even lead to them coordinating to one another to form so-called "pi-dimers" or "pi-stacks", rather than adopting more conventional intercalated salt structures.

Experimental Details

As part of our ongoing studies on TAC ionic liquids, we have prepared tris(ethylmethylamino)cyclopropenium chloride (**1**) and tris(diethylamino)cyclopropenium iodide (**2**). Salt **1** was prepared by the reaction of C_3Cl_5H with NEtMeH, whereas salt **2** was prepared by reaction of $[C_3(NEt_2)_3]Cl$ with EtI. Bielawski has reported similar reactions of halide salts with alkylating agents, such as Me_2SO_4, MeOTs and $[R_3O]BF_4$,[18] however, we believe this is the first synthesis of an iodide salt by reaction of a chloride with

EtI. Ion chromatography confirmed that the chloride had been replaced very efficiently by this route. Single crystals of these materials slowly formed in the neat liquids. Milyukov recently reported the synthesis of **2** via metathesis of the chloride salt with KI.[19] They also reported its room temperature crystal structure.

Results and Discussion

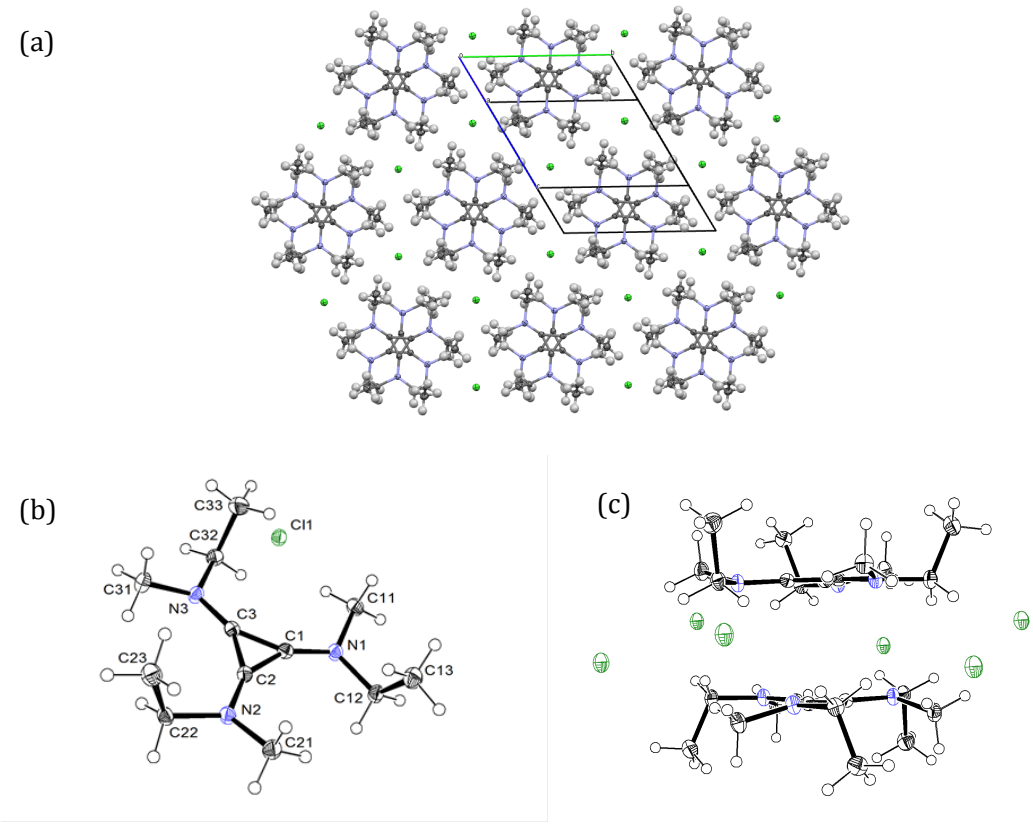


Figure 1. (a) Hexagonal layering of **1**, (b) The cation of **1** with the atomic numbering scheme, (c) side-on view of the dicationic dimer unit with surrounding counterions. Atoms are represented as 40% probability thermal ellipsoids.

Crystal structures

The 120 K P-1 solid state structure of **1** reveals *solvent-free* close-packed layers of triaminocyclopropenium dimers with chloride anions at the hexagonal vertices within the layers (Figure 1a). The most noteworthy aspect of the structure is the short distance between the cations (Figure 1c), with the cyclopropenium C_3 centroids separated by only 3.2251(13) Å. This means the

two positively-charged aromatics are closer than in most “pi-stacked” neutral arenes at 3.3–3.8 Å.[20–25] The interlayer distance in graphite, to give another useful comparison, is 3.35 Å.

It is also worth noting that the anions appear to be associated with the alkyl CH groups rather than the formally cationic C_3N_3 core, although this is not without precedent in the literature. The solid state structure of unsolvated $[C_3(NCy_2)_2(NEt_2)]Cl$ was recently reported, also with Cl^- ions in close proximity to the alkyl groups.[26] In this case, it appears the bulky amino groups prevent dimer formation.

The view down the intermolecular C_3–C_3 centroid-centroid axis indicates that each Me group lies over an Et group. Steric repulsion between the protons on these groups appears to push the Me groups out of the plane more than the bulkier Et groups which would be interacting more with the layers above and below.

The 120 K solid state structure of **2** was also determined to explore the effect of changing alkyl substituents and counterions (Figure 2). The low temperature $[C_3(NEt_2)_3]I$ structure was solved in the same space group as the previous room temperature structure[19] ($C_{2/c}$) with two cation/anion pairs per unit cell. Despite the different space groups of **1** and **2**, the packing is quite similar to the chloride salt described above, with the extended structure also consisting of TAC sandwich dimers in a close-packed layer arrangement with the iodide anions within these layers. Again, the dimers are in a staggered conformation, and in close proximity to one another with a C_3–C_3 distance of 3.351(2) Å.

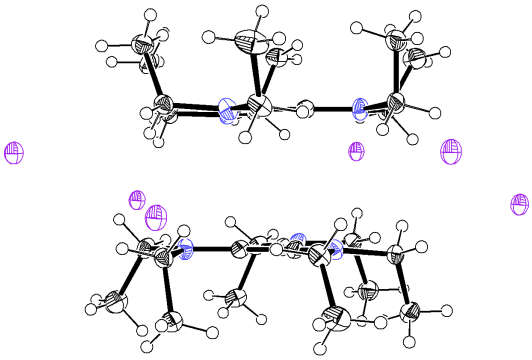


Figure 2. Side-on view of the dicationic dimer unit with surrounding counterions for compound **2**. Atoms are represented as 40% probability thermal ellipsoids.

A detailed side-by-side comparison of structural parameters for compounds **1** and **2** is presented in Table 1. These observations raise some interesting questions: Why does the dicationic dimer form in preference to more conventional structures with alternating cations and anions? Why is the intermonomer distance so short? What are the fundamental interactions responsible for stabilizing the dicationic sandwich motif?

Computational Analysis

To address these questions, interaction energy curves were calculated for the dissociating dimer of compound **1** as an isolated system in the gas phase, and with the crystalline environment modelled using explicit chloride counterions with a partial nuclear charge model used to ensure overall charge balance while maintaining point-group symmetry in the system (Z_{eff} = 17.666). The three closest counterions were assigned to each monomer unit (Figure 1c) and displaced accordingly upon dissociation. Gas phase potential energy curves were evaluated at MP2/jun-cc-pVDZ and SAPT0/jun-cc-pVDZ while counterion charge-balanced energies were evaluated at MP2/jun-cc-pVDZ only. MP2 energies were corrected for basis set superposition error throughout. Local well depths for metastable complexes were recalculated at SAPT0/aug-cc-pVDZ and SAPT0/jun-cc-pVTZ. Further details of the computational methods and references are available in Appendix 2. All interaction energies are reported relative to the infinitely separated monomer units, and all displacements relative to the experimental intermolecular C_3 - C_3 centroid distance for **1**.

Table 1. Observations on the crystal structures of **1** and **2**.

Parameter*	(**1**)	(**2**)	Observation/Comment
$r_{C_3–C_3}$	3.2251(13) Å	3.351(2) Å	Remarkably short relative even to neutral pi-stacked dimers
$r_{N_3–N_3}$	3.1195(13) Å	3.287(2) Å	Nitrogen atoms bent slightly toward centre of dimer
$r_{Me_3–Me_3}$ φC-C-N-Me	3.8356(13) Å 21.9°	– –	Methyl groups bent significantly away from centre of dimer (**1**)
$r_{Et_3–Et_3}$ φC-C-N-Et	3.1099(13) Å –4.9°	3.4380(18) Å 6.6°	Ethyl groups either bent slightly towards the centre of the dimer (**1**) or slightly away from the centre of the dimer (**2**)
$r_{C–C}$	1.384(2) Å	1.383(3) Å	Relatively short C-C distances reflect aromatic nature of TAC^+
$r_{C–N}$	1.329(4) Å	1.330(5) Å	Relatively short C-N distances confirm delocalized aromatic system extends out to N atoms
$\{r_{Cl–Me}$, $r_{Cl–Et}$, $r_{Cl–Et\beta}\}$	3.56–4.31 Å	3.94–4.53 Å	Relatively long halide–ethyl/methyl carbon distances indicate only weak long-range electrostatic interactions may be present

*C denotes the cyclopropenium carbons, Me denotes the methyl carbons, Et denotes the methylene carbons of the ethyl substituent, Etβ denotes the terminal carbons of the ethyl substituent. The 3 subscript denotes the corresponding centroid of all symmetry equivalent atoms within the molecule.

To investigate the energetics of dicationic dimer formation, it is necessary to initially exclude long range electrostatic stabilization effects. Therefore, isolated dimer interaction energy curves are presented in Figure 3.

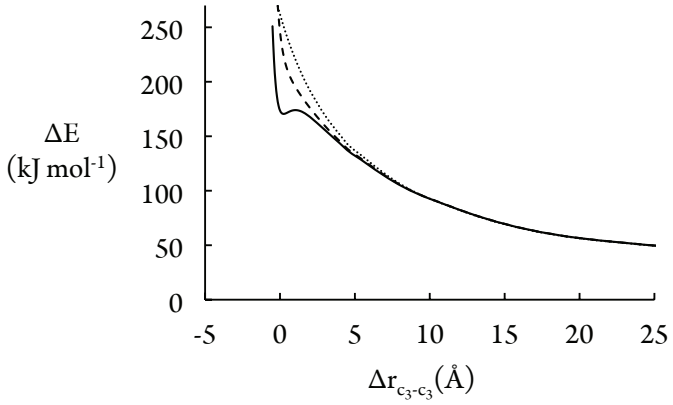


Figure 3. BSSE-corrected HF/jun-cc-pVDZ (dashed line), SAPT0/jun-cc-pVDZ and MP2/jun-cc-pVDZ (solid line, indistinguishable) interaction energy curves, and extrapolated 1/r asymptotic electrostatic potential energy curve (dotted line).

These interaction energy curves reveal that intermolecular electron correlation, *i.e.* dispersion, is the primary force driving dimerization *at short range*. Only

methods that account for intermonomer electron correlation predict transient metastable complex formation, which can then be 'locked in' by long-range electrostatic interactions with the surrounding crystalline matrix. Conversely, if dispersion forces were unable to overcome the 'baseline' electrostatic repulsion to create a metastable dicationic dimer complex, then this motif would not occur within the crystalline environment at all, i.e. some other packing arrangement would be observed.

It is widely acknowledged that modelling crystal packing forces and predicting solid state structures from gas phase interaction potentials is a complex and challenging problem.[27] However, all modern approaches to solving this problem agree that proximal environmental effects are best captured using screened or fuzzy charge models,[28-32] while longer range electrostatic interactions can be well described using classical point charge or polarizable dipole models.[32]

In this spirit, we employ 6 chloride counterions with fictitious nuclear charges of 17.666 such that each counterion overall carries a partial negative charge of 0.333, as the simplest possible approach that is consistent with obtaining the correct asymptotic electrostatic behaviour of the surrounding environment while maintaining a realistic density-based representation of the counterions closest to each dicationic dimer unit.

Interaction energy curves for the dissociation of this 'charge balanced' system along the $C_3 - C_3$ centroid axis are illustrated in Figure 4.

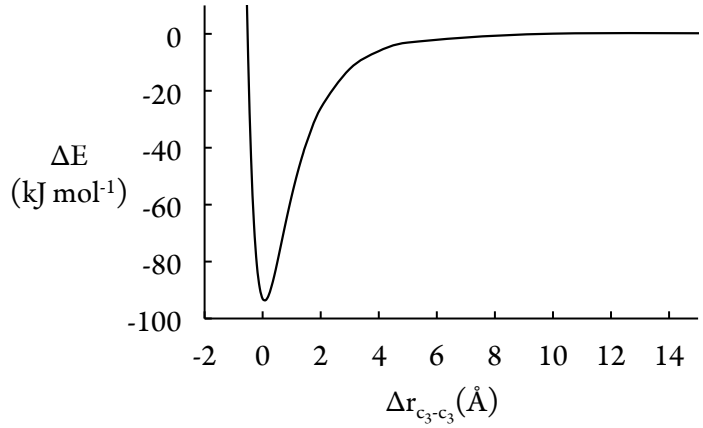


Figure 4. BSSE-corrected MP2/jun-cc-pVDZ interaction potential for dicationic dimer dissociation, with three partially charged Cl^- ions surrounding each monomer unit. The predicted minimum lies within 0.07 Å of the observed C_3-C_3 centroid distance, validating our partial charge model for the crystalline environment.

The stabilization enthalpy of -93.7 kJ mol^{-1} can only be considered an extreme lower bound to the experimental enthalpy of formation, given the approximate treatment of environmental interactions. Nonetheless, these results clearly demonstrate the well-known contribution that long-range electrostatic interactions make to the high thermal stability of these compounds.

Closer inspection of Figures 3 and 4 reveals that the position of the metastable local minimum in the gas phase is slightly longer than in the condensed phase, at distances of 3.47 Å and 3.30 Å, respectively. This difference can be attributed to the chloride counterions forming a weak electrostatic bridge at short range, in addition to providing long-range electrostatic stabilization. This enhanced interaction between monomer units leads to the remarkably short experimentally-observed intermolecular C_3 centroid distance. The iodide ions in complex **2** form a weaker electrostatic bridge, due to their larger ionic radii and lower electronegativities, resulting in the longer experimentally observed intermonomer separation.

Overall, these results are consistent with previous computational studies on dicationic and dianionic dimers formed from charged organic radical species, which concluded that dispersion was the dominant short-range stabilizing force

in these systems, with intermolecular π-bond formation playing a secondary role.[33-40] For the TAC dimer dication, the story is even neater, as the closed-shell nature of the monomer units prohibits intermolecular bond formation, further evidenced by the purely repulsive HF curve and molecular orbital analysis (Appendix 2). *Therefore, dispersion forces primarily drive dicationic dimer formation in these systems*, while the surrounding counterions form an additional weak electrostatic bridge between the monomer units at short range, but more importantly provide a long range background source of charge balance to stabilize each dicationic dimer within the crystalline environment. Without these attractive dispersion interactions between monomer units, crystal packing forces would likely lead to the formation of more conventional structures with alternating cations and anions.

Finally, it remains to consider the electrostatic properties of the monomer units. Existing evidence suggests that the positive charge is delocalized over the TAC scaffold.[8,9,13–17] Atom-centred density partitioning and electrostatic potential fitting analyses of the HF density (Appendix 2) agree that the positive charge is delocalized over the carbon and hydrogen atoms, while the nitrogen atoms carry partial negative charges. The ethyl and methyl substituents carry a much larger share of the positive unit charge than the inner cyclopropenium ring. This implies that dicationic dimer formation depends on the ability of each monomer to 'hide' positive charge away from the dimer interface.

Table 2. Local well depths (kJ mol^{-1}) in SAPT0 dissociation energy curves for gas phase dicationic dimers.

Amino groups	jun-cc-pVDZ	aug-cc-pVDZ	jun-cc-pVTZ
NEt2 (2)	6.7	14.1	14.8
NEtMe (1)	5.3	12.2	12.9
NMe2	4.3	11.3	12.1
NH2	No local well, repulsive potential		

Local well depths for metastable dimers with varying sizes of amino groups are presented in Table 2. Dimer formation becomes more favourable with alkyl

substitution and increasing alkyl chain length, as the positive charge becomes increasingly delocalized.

The 60° rotational offset around the inter-monomer axis observed experimentally also serves to minimize the electrostatic repulsion between monomer units and allow transient dimer formation, as illustrated by the aligned and rotationally offset interaction energy curves in Figure 5.

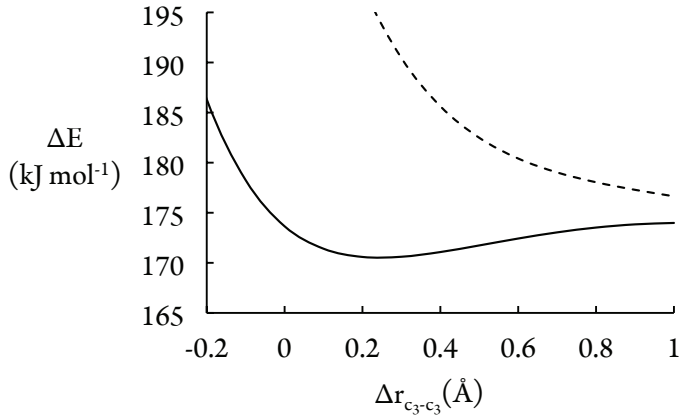


Figure 5. BSSE-corrected MP2/jun-cc-pVDZ interaction energy curves for the aligned dimer **1** (dashed line) and the rotationally-offset conformation (the solid line represents a subset of data from Figure 3).

Conclusion

In summary, these remarkable systems are paradigm-shifting in a number of ways: they represent the first known example of *unsolvated* TDAC halides, in which the cationic cyclopropenium monomer units preferentially coordinate to one another rather than the surrounding counterions or solvent molecules. Even more remarkably, the monomers are found *closer together than usually observed in 'pi-stacked' systems of neutral aromatics*. Counter-intuitively, dicationic dimer formation is mainly driven by *strong dispersion interactions* supplemented by *relatively weak electrostatic interactions with counterions at short range*, and consolidated by long-range electrostatic interactions that charge balance the system. This novel observation turns the textbook understanding of intermolecular interactions on its head.

Supporting Information

Appendix 2 contains full experimental details concerning the synthesis, crystallisation, characterisation and crystallographic data; computational methods, optimized molecular coordinates of dicationic dimer surrounded by chloride counterions with scaled nuclear charges, and *ab initio* data including absolute energies in Hartrees underlying all Figures, molecular orbital plots and atomic partial charges. Full crystallographic data is also available from the Cambridge Crystallographic Data Centre (CCDC 1062303 and 1062302).

References

(1) Wasserscheid, P. & Welton, T. *Ionic Liquids in Synthesis*, 2nd Ed Wiley-VCH: Weinheim, 2008.

(2) Saielli, G., Bagno, A. & Wang, Y. *J. Phys. Chem. B* **2015**, *119*, 3829–3836.

(3) Urahata, S. M. & Ribeiro, M. C. C. *J. Chem. Phys.* **2004**, *120*, 1855–1863.

(4) Wang, Y. & Voth, G. A. *J. Am. Chem. Soc.* **2005**, *127*, 12192–12193.

(5) Canongia Lopes, J. N. A. & Pádua, A. A. H. *J. Phys. Chem. B* **2006**, *110*, 3330–3335.

(6) Triolo, A., Russina, O., Bleif, H. & Di Cola, E. *J. Phys. Chem. B* **2007,** *111*, 4641–4644.

(7) Curnow, O. J., MacFarlane, D. R. & Walst, K. J. *Chem. Commun.* **2011**, *47*, 10248–10250.

(8) Johnson, R. W. *Tetrahedron Lett.* **1976**, *17*, 589–592.

(9) Jiang, Y., Freyer, J. L., Cotanda, P., Brucks, S. D., Killops, K. L., Bandar, J. S., Torsitano, C., Balsara, N. P., Lambert T. H. & Campos, L. M. *Nature Comm.* **2015**, *6*, 6950

(10) Komatsu K. & Kitagawa T. *Chem. Rev.* **2003,** *103*, 1371–1427.

(11) Bandar, J. S. & Lambert, T. H. *Synthesis* **2013**, *45*, 2485–2498.

(12) Yoshida, Z. I.; Tawara, Y. *J. Am. Chem. Soc.* **1971**, *93*, 2573–2574.

(13) Weiss R., Brenner T., Hampel F. & Wolski A. *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 439–441.

(14) Butchard, J. R., Curnow, O. J., Garrett, D. J. & Maclagan, R. G. *Angew. Chem.* **2006**, *118*, 7712–7715.

(15) Butchard, J. R., Curnow, O. J., Garrett, D. J., Maclagan, R. G., Libowitzky, E., Piccoli, P. M. & Schultz, A. J. *Dalton Trans.* **2012**, *41*, 11765–11775.

(16) Weiss, R., Rechinger, M., Hampel, F. & Wolski, A. *Angew. Chemie, Int. Ed.* **1995**, *34*, 441–443.

(17) Weiss, R., Schwab, O. & Hampel, F. *Chem. Eur. J.* **1999**, *5*, 968–974.

(18) Vu, P. D., Boydston, A. J. & Bielawski, C. W. *Green Chem.* **2007**, *9*, 1158–1159.

(19) Milyukov, V. A., Shakirova, L. R., Bezkishko, I. A., Krivolapov, D. B. & Sinyashin O. G. *Russ. Chem. Bull, Int. Ed.* **2012**, *61*, 1483–1488.

(20) Janiak, C. *J. Chem. Soc., Dalton Trans.* **2000**, 3885–3896.

(21) Martinez, C. R. & Iverson, B. L. *Chem. Sci.* **2012**, *3*, 2191–2201.

(22) Hunter, C. A. & Sanders J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525–5534.

(23) Amabilino, D. B. & Stoddart, J. F. *Chem. Rev.* **1995**, *95*, 2725–2829.

(24) Ringer, A. L. & Sherrill, C. D. *J. Am. Chem. Soc.* **2009**, *131*, 4574–4575.

(25) Wheeler, S. E. & Houk, K. N. *J. Am. Chem. Soc.* **2008**, *130*, 10854–10855.

(26) Bandar, J. S., Tanaset, A. & Lambert, T. H. *Chem. Eur. J.* **2015,** *21,* 7365–7368.

(27) Price, S. L. *Chem. Soc. Rev.* **2014,** *43,* 2098–2111.

(28) Wang, B. & Truhlar, D. G. *J. Chem. Theor. Comput..* **2012,** *8,* 1989–1998.

(29) Wang, B & Truhlar, D. G. *J. Chem. Theor. Comput..* **2014,** *10,* 4480–4487.

(30) Wang, Q. T., Rackers, J. A., He, C., Qi, R., Narth, C., Lagardere, L., Gresh, N., Ponder, J. W., Piquemal, J. P. & Ren, P. Y. *J. Chem. Theor. Comput..* **2015,** *11,* 2609–2618.

(31) Besalu, E. & Carbo-Dorca, R. *J. Math. Chem.* **2013,** *51,* 1772–1783.

(32) Olsen, J. M. H., Steinmann, C., Ruud, K. & Kongsted, J. *J. Phys. Chem. A* **2015,** *119,* 5344–5355.

(33) Garcia-Yoldi, I., Miller, J. S. & Novoa, J. J. *J. Phys. Chem. A* **2009**, *113*, 484–492.

(34) Lu, J. M., Rosokha, S. V. & Kochi, J. K. *J. Am. Chem. Soc.* **2003**, *125*, 12161–12171.

(35) Jakowski, J. & Simons, J. *J. Am. Chem. Soc.* **2003**, *125*, 16089–16096.

(36) Jung, Y. & Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2008–2011.

(37) Braida, B., Hendrickx, K., Domin, D., Dinnocenzo, J. P. & Hiberty, P. C. *J. Chem. Theory Comput.* **2013**, *9*, 2276–2285.

(38) Novoa, J. J., Lafuente, P., Del Sesto, R. E. & Miller, J. S. *Angew. Chem. Int. Ed.* **2001**, *40*, 2540–2545.

(39) Del Sesto, R. E., Miller, J. S., Lafuente, P. & Novoa, J. J. *Chem. Eur. J*. **2002**, *8*, 4894–4908.

(40) Mota, F., Miller, J. S. & Novoa, J. J. *J. Am. Chem. Soc.* **2009**, *131*, 7699–7707.

4. An Equation-of-Motion Coupled-Cluster Study of the Electronic Spectrum of Zinc Phthalocyanine

Abstract:

The metallophthalocyanines are widely used as dyes, in molecular electronics, and as photo-sensitizers due to their optical and electronic properties. Yet despite 30 years of computational analysis, assignments of the transitions in their electronic absorption spectra remain uncertain.

Here we examine the electronic excitation spectrum of zinc phthalocyanine (ZnPc). ZnPc is one of the simplest of the metallophthalocyanines, as it is a closed shell system and all *3d*-orbitals on the metal ion are fully occupied. Recent studies using TDDFT methods have failed to reproduce the experimentally observed number of peaks in each region of the absorption spectrum, and the assignment of transitions beyond the first excited state of each symmetry remains uncertain. Here we show that the equation-of-motion coupled-cluster (EOM-CC) methods can identify the correct number of peaks. Calculated transition energies are correct to within an average absolute error of 0.26 eV at EOM-CCSD(T)/3-21G*, with correctly ordered transition intensities.

Introduction:

The phthalocyanine (Pc) family of molecules (Figure 1) are organic chromophores, that demonstrate high absorptivity within the visible region due to spatial overlap between the π and π* orbitals in the extended aromatic ring system. Phthalocyanines are among the most thermally stable of any organic molecules, subliming at around 500°C. As a result, metal phthalocyanines (MPcs) and their derivatives are widely used as dyes and pigments, accounting for 25% of all organic pigment use worldwide.[1] Their optical and electronic properties, as well as their chemical similarity to chlorophyll and heme, have also led to interest in their electronic structure and their potential use in a variety of areas

including solar cells,[2-6] organic light-emitting diodes (OLEDs),[6-8] and photosensitizers for photodynamic therapy (PDT) for cancer treatment.[9,10]

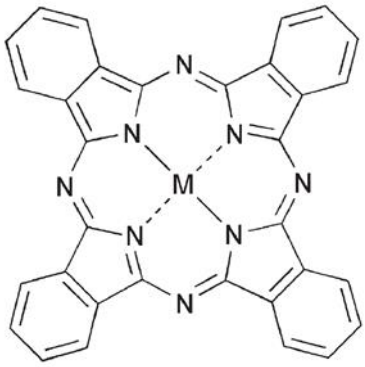


Figure 1: The structure of transition metal phthalocyanines. The phthalocyanine ligand is planar, with D_{4h} symmetry. It contains eight nitrogen atoms in the macrocycle surrounding the central metal ion.

Phthalocyanine spectra are dominated by a multitude of strong ligand π -> π* transitions, that give rise to complex spectra with a significant number of transitions within the visible region of the spectrum. Possible excitations from the lone pairs of the nitrogen atoms into the π* orbitals, and potential transitions to Rydberg states have also been discussed in the literature. Phthalocyanine spectra are characterized by a sharp peak near 2 eV (~650nm), known as the Q-band, and a much broader and much more intense band around 3 eV (~400nm), known as the B-band, which gives many phthalocyanines their intense deep-blue colouring. The metal ions Fe and Mn introduce metal *d*-orbitals that sit between the ligand HOMO and LUMO in energy, and the subsequent charge-transfer transitions between the metal and ligand add additional complexity to the spectrum, making MnPc black in colour as it has significant absorption across almost the entire visible range, while CuPc is a deep royal blue and is used for the blue ink in most colour printers.[11]

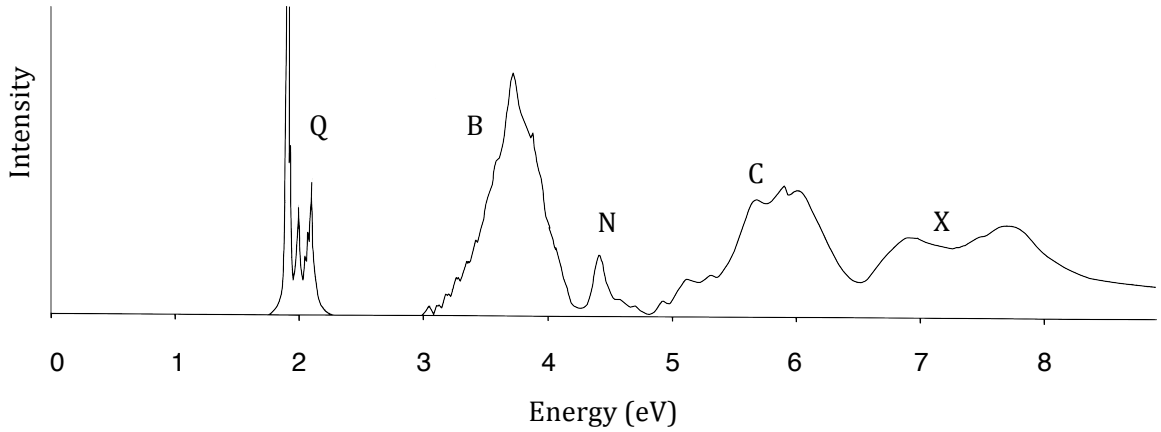


Figure 2: The absorption spectra of ZnPc in an argon matrix at 5K from images in Vancott *et al*[12] with the traditional labelling of the absorption bands added.

ZnPc Spectra

UV-Vis spectra of ZnPc have been collected in a variety of environments, including thin film,[13-15] gas phase,[16-18] in solution,[19-22] and in an argon matrix.[12,21] These spectra are similar, indicating that the environment does not tend to substantially influence the energies or intensities of the peaks for this system. However, the spectra taken at higher temperatures have significantly broader peaks, and the thin film spectra show Davydov splitting. The sharpest spectrum is that obtained by Vancott *et al.*[12] in an argon matrix at low temperature, and the spectrum shown in Figure 2 is reconstructed from the spectral images provided therein. That spectrum is sufficiently sharp to show vibronic structure (the three vibronic peaks of the Q band are clear in Figure 2, as is the structure on the left side of the B band). Since the early work by Gouterman on these systems in the 60s it has been traditional to refer to the absorption bands with letters,[23,24] and these have been included in Figure 2. Table 1 gives numerical data on the range of these bands.

Vancott *et al.* report the locations of the peaks and intensities in their spectrum, and those values are reproduced in Table 2. Some bands are very broad (e.g. B), which leave open the question of where exactly in the band the electronic transitions actually lie. In cases of overlapping peaks, assigning empirically observed oscillator strengths can also become somewhat arbitrary, as it can be

difficult to know how much each peak within the overlapping region contributes to the observed intensity.

Table 1: Standard labels for the band regions used in the literature, and the approximate range of each band.

Band Name	Energy range (eV)
Q	1.8 - 2.1
B	3.0 - 4.2
N	4.3 - 4.8
L	4.8 - 5.5
C	5.5 - 6.5
X	6.6 - 8.7

Table 2: Peaks in the argon matrix spectrum as reported by Vancott *et al.*[12] Oscillator strengths calculated from data provided therein.

Peak Name	Transition energy (eV)	Oscillator strength *f*
Q	1.89	0.40
Q'	2.08	0.03
B1	3.74†	1.25
B2	3.71	0.04
B3*	3.99	0.00
N1	4.42	0.11
N2	4.70	0.02
L1	4.88	0.01
L2	5.10	0.10
L3	5.33	0.10
C1	5.62	0.30
C2	5.92	0.03
C3	6.00	0.85
X1	6.89	0.58
X2	7.67	0.81

* They were able to determine that peak B3 has a different symmetry from other strongly allowed transitions in that range, as its intensity changed with the angle of applied light in a way that the nearby peaks did not.
† This is a very broad band. The stated transition energy is the centre of the band.

Previous computational work

The relatively large size of the phthalocyanine molecules (56 ligand atoms) limits the accuracy of computational models that can be applied to model its electronic structure and spectrum. In the 1960s and 70s, Gouterman and coworkers pioneered the use of configuration interaction methods on the phthalocyanines, but could only use 4 orbitals in the active space (HOMO, HOMO-1, and the degenerate LUMO orbitals).[23,24] This describes only the transition from

the ground state to the first excited state (the 'Q' transition). More recently, time-dependent density functional (TDDFT) techniques have been applied to this problem, with dozens of papers published in the last 15 years. However, different density functionals predict qualitatively different ground states for manganese phthalocyanine and iron phthalocyanine.[25,26] TDDFT methods have failed to predict the empirically observed number of transitions within the spectral bands.[25,27] As a result, a number of authors have called for the application of advanced *ab initio* multireference techniques to these systems.[12,25,28]

Some of the metallophthalocyanines have overlapping peaks, and so measuring agreement between calculated and observed spectra can be difficult. Zinc phthalocyanine (ZnPc) provides a useful benchmark system, as it has a closed d-shell and so there are no charge transfer transitions between the metal and the ligand in the spectrum, which means it has one of the simplest spectra of the MPc family. As the charge on the central metal ion changes across the period, the energies and occupancies of the metal *3d* orbitals change, sometimes falling within the ligand's HOMO - LUMO gap, and complicating the spectra of the other MPcs with transitions to and from these orbitals.[26] By contrast, ZnPc has a sharp first excited state peak (called 'Q') that is well-separated from other peaks. It is universally agreed that this corresponds to a $\pi \rightarrow \pi^*$ transition from the ligand HOMO to the ligand LUMO, which allows for a clear measure of the margin of error in the calculations. Despite the relative simplicity of the ZnPc spectrum, there is no agreement in the literature about the assignments of any excited states beyond the first peak of each symmetry.

On the ZnPc system itself, two main types of quantum mechanical calculations have been carried out on this system in the last 40 years. Semi-empirical calculations were published by the Mack and Stillman group in a series of papers in the 1990s and early 2000s[29-32] of which their 2001 summary is representative.[31] They make use of Zerner's intermediate neglect of differential overlap method (ZINDO),[33] which is a variant of Pople's intermediate neglect of differential overlap method (INDO).[34] These methods use a minimal basis set and

discard many integrals in the calculations for efficiency. These approximations lead to very inaccurate energies, and so empirical data from the molecule is entered as parameters for the method in order to yield more accurate computational results. Since the development of TDDFT methods, TDDFT calculations have been carried out on ZnPc by a series of authors since 2001, mostly used the B3LYP functional,[28,35-38] of which Ueno's 2012 paper is representative.[28] Table 3 matches previous ZINDO and TDDFT results against experimental results based on symmetry, transition energies, and intensities.

Table 3 Literature TDDFT and ZINDO calculated transition energies (eV) and oscillator strengths (*f*).

	Experiment[12]		B3LYP/ 6-311+G(*d*,*p*)[28]		M11/ 6-31G*[39]		SAOP/ STO-TZ+pol[40]		ZINDO[31]†	
Peak Name	eV	*f*	eV	*f*	eV	*f*	eV	*f*	eV	*f*
Q	1.89	0.40	2.05	0.43	1.88	0.53	1.96	0.74	1.83	0.90
Q'	2.08	0.03			2.10	0.44				
							2.87	0.03		
							3.07	0.05		
			3.35	0.01			3.14	0.30		
			3.64	0.16			3.28*	0.00		
			3.81	0.03			3.34	0.04		
B2	3.71	0.04	3.71	0.28	3.29	0.10	3.50	0.66	3.72	0.02
B1	3.74	1.25	3.85	0.50	3.82	0.09	3.81	1.15	4.30	2.23
B3	3.99*	0.00	3.93*	0.00			4.23*	0.00	4.23*	0.03
							4.31*	0.00		
							4.32	0.03		
							4.50	0.01		
							4.50	0.02		
							4.58	0.04		
							4.64	0.05		
							4.69	0.02		
N1	4.42	0.11	4.30	0.20			4.77	0.14	4.25	0.43
							4.78*	0.00		
							4.81	0.00		
N2	4.70	0.02					4.98	0.02	4.53	0.33
L1	4.88	0.01							4.83	0.01
L2	5.10	0.10							5.10	0.02
L3	5.33	0.10								
C1	5.62	0.30								
C2	5.92	0.03								
C3	6.00	0.85								
X1	6.89	0.58								
X2	7.67	0.81								

* A state with different symmetry to the other states according to the authors
† The authors themselves assign the calculated peaks to different transitions than the assignments made here

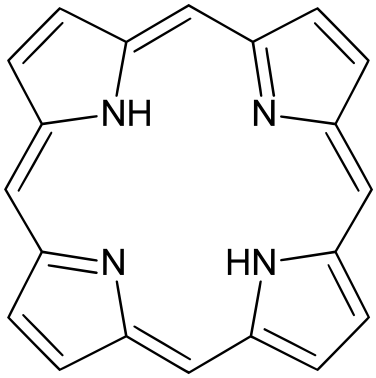


Figure 3: Free base porphin. Porphyrins are similar to the phthalocyanines but lack the stability provided by the fused benzene rings and the N-substitutions on the macrocycle. Like the phthalocyanines, porphyrins often have a central metal, although in free base porphin the metal is absent and is replaced by two H atoms, which reduce the symmetry from D_{4h} to D_{2h}.

Previous studies on free-base porphin (Figure 3) provide an indication of the likely utility of different computational methods for modelling the electronic spectrum of ZnPc. The porphyrin systems have about half the atoms of the phthalocyanines, but have the same general shape, allowing them to be used as a first-degree approximation as to how different computational methods would be likely to perform on a phthalocyanine-like macrocyclic ligand system. Chaudhuri *et al.*[41] provide a survey of the results of computational studies on the spectrum of free-base porphin, observing that EOM-CCSD calculations reproduce the spectrum with errors below 0.3 eV for every peak,[41] a slightly larger error margin than is usually seen from EOM-CCSD (<0.2 eV).[42,43] They found active space multireference methods with a correlation correction (e.g. MRMP, CASPT2, NEVPT2, IVO-MRMP) had a similar margin of error to the EOM-CCSD method for this system. However, CI methods without a perturbation theory correction (such as CASSCF and IVO-CASCI) had larger errors in energy - ranging from 0.5 eV to 1.6 eV. These inaccuracies were systemic, with the ground state always being energetically favoured over the excited states, yielding excitation energies that were consistently too high.

Methods

Calculations of EOM-CCSD excited state energies and transition moments in this chapter were performed with the GAMESS,[44,45] NWChem,[46] ACES II,[47] and

QChem[48] computational chemistry packages. The geometry used for zinc phthalocyanine is given in Appendix 3, as is the absolute energies corresponding to the calculation results given in this chapter.

The algorithms used by these programs for calculating the non-iterative triples corrections vary widely and are described in a number of papers.[43,49-59] Table 4 lists the algorithms used here from these programs. With the exception of NWChem, all software packages used offered multiple perturbative triples algorithms. Within GAMESS we used the two methods that the developers describe in the documentation as the most accurate. The CR-EOML (2,3),D method is reported to recover the most accurate energies for the ground and excited states individually, while the DEL(IID) correction is designed to be more accurate for transition energies, as it corrects for the fact that the ground state is better described than excited states after the triples correction has been applied to each. Although multiple perturbative triples methods are available in AcesII, only one ran successfully on this system, CCSD-1. We chose simplest of the methods QChem offers, for computational efficiency.

Table 4: Triples-corrected, EOM-CCSD(T)-like, algorithms used in this work, loosely ordered from theoretically least rigorous on the left, to theoretically most-rigorous on the right.

QChem	AcesII	GAMESS	NWChem	GAMESS
CCSD(fT) aka EOM-CCSD(2)$_T$[59]	CCSDT-1[52] Similar to CCSD(TQ)	CR-EOML (2,3),D[54] aka CR-EOM-CCSD(T)$_L$ or CR-EOM-CC(2,3)	CR-EOM-CCSD(T)[51]	CR-EOML, DEL(IID)[53,60] aka δ_{IID}-CR-EOM-CC(2,3)

When performing non-iterative triples corrections, it was necessary to freeze some orbitals in order to reduce the memory requirements and run-time. Freezing orbitals also became necessary for EOM-CCSD calculations when using basis sets larger than 3-21G. The chemical core orbitals were frozen in all calculations reported in this chapter.

Zinc Phthalocyanine belongs to the D_{4h} point group, although D_{2h} is the largest Abelian subgroup, and thus calculations were run in D_{2h} symmetry. This work will use the D_{2h} symmetry labels for orbitals and states.

Results and discussion

EOM-CCSD/3-21G results shown in Table 5 and Figure 4 reproduce trends in experimental transition energies, apart from a constant systemic error of around +0.6 eV relative to the ground state energy. Once the constant offset error is accounted for, most transition energies are within 0.2 eV of experimental values. An exception to this is the highest energy transitions that are in error by up to 0.7 eV even after the systemic offset is applied. This difference appears likely due to the fact that all transitions below the C-band, except L1, are characterized predominantly by transitions from occupied orbitals into the LUMO orbital, while the transitions in the higher energy C-band are all characterized predominantly by transitions to higher energy virtual orbitals, which may have required a larger basis set to accurately describe. The calculated oscillator strengths are qualitatively in agreement with the experimental values, and provide an alternate way of confirming state assignments.

The ground state of ZnPc is well established as being a singlet A_g state. Symmetry allowed excitations are to B_{2u}/B_{3u} states (which are degenerate), and to B_{1u} states. Our calculations show that of the first 12 allowed transitions in the spectrum, 12 are to states of B_{2u}/B_{3u} symmetry while 2 are to states of B_{1u} symmetry. In Table 5, the two transitions to B_{1u} states are flagged, and both have low oscillator strengths of 0.01. The remainder of the transitions are to B_{2u}/B_{3u} states. The assignment of the first B_{1u} transition to the B3 peak is confirmed by Vancott *et al.* who were able to deduce experimentally that the symmetry of state associated with the B3 peak was different to nearby peaks due to its intensity changing with the angle of applied light.[12]

Table 5: EOM-CCSD/3-21G energies and dimensionless oscillator strengths (*f*), of symmetry-allowed B_{1u} and B_{2u}/B_{3u} states. All states are degenerate B_{2u}/B_{3u} states unless otherwise marked.

Experimental			Calculation		Error (eV)
Name	Energy (eV)	f	Energy (eV)	f	
Q	1.89	0.40	2.42	0.41	0.53
B1	3.74	1.25	4.34	1.07	0.6
B2	3.71	0.04	4.52	0.22	0.81
B3	3.99	0.00	4.57 (B_{1u})	0.01	0.58
N1	4.42	0.11	4.82	0.40	0.4
N2	4.70	0.02	5.06	0.15	0.36
L1	4.88	0.01	5.30	0.10	0.42
L2	5.10	0.10	5.70	0.18	0.6
L3	5.33	0.10	6.16	0.11	0.83
C1	5.62	0.30	6.85 (B_{1u})	0.01	1.23
C2	5.92	0.03	6.95	0.05	1.03
C3	6.00	0.85	7.00	0.68	1

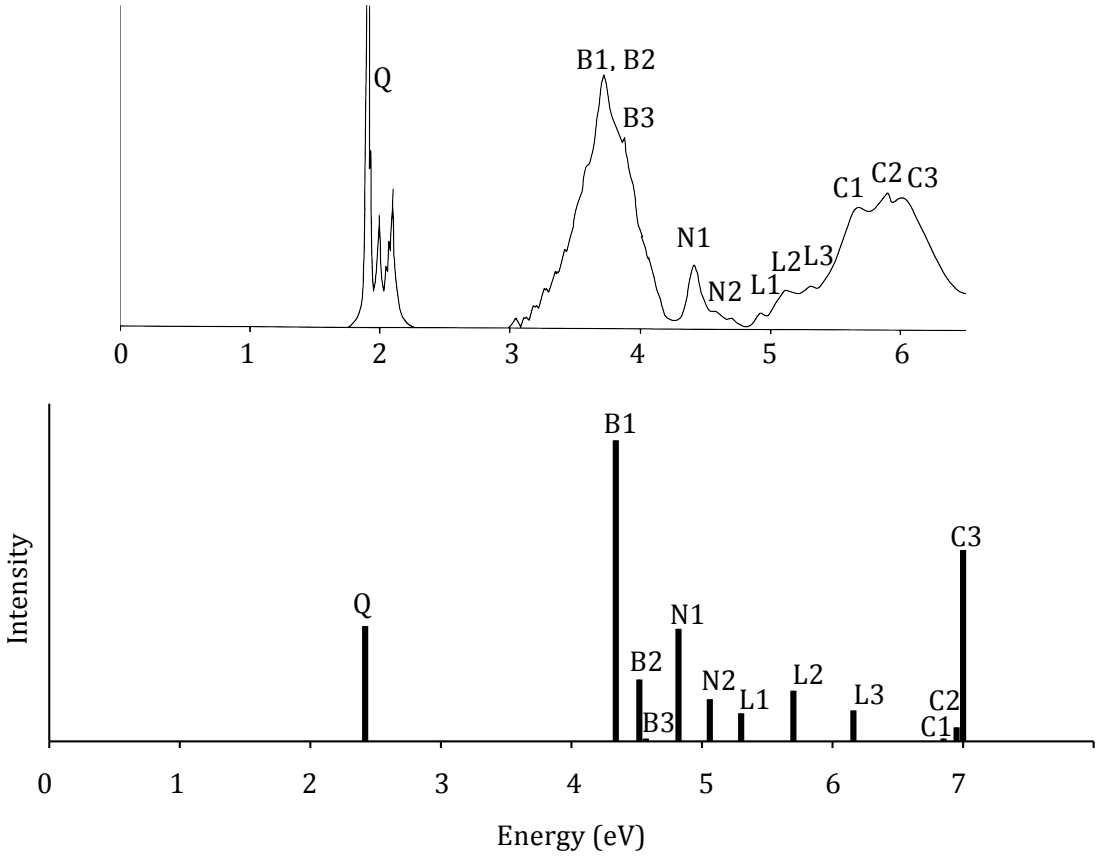


Figure 4: The experimental spectrum with the calculated EOM-CCSD/3-21G spectrum. Spectra are offset from each other in order to align the first peak (Q).

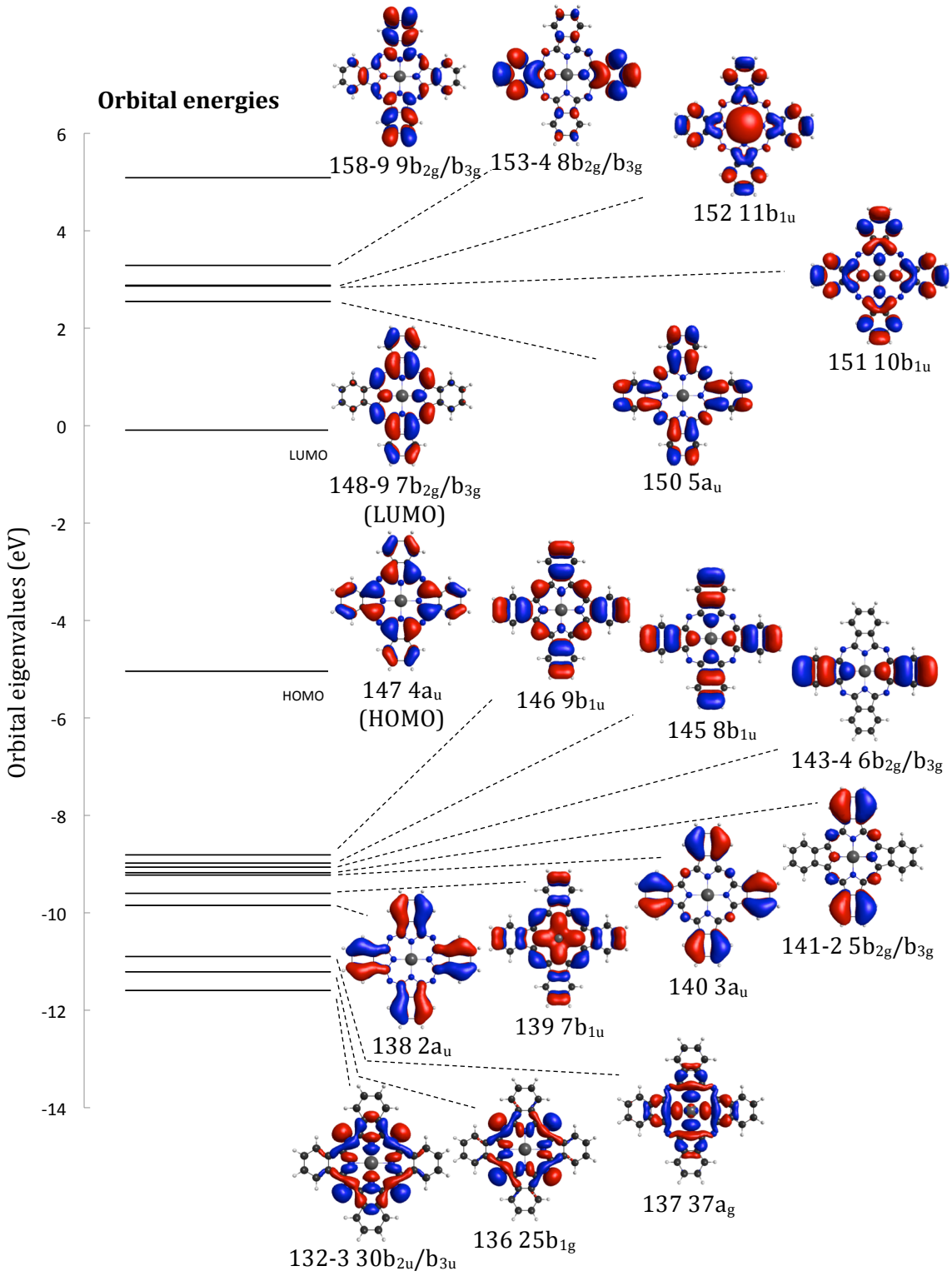


Figure 5: Molecular orbitals and energies (eigenvalues) of the ZnPc HF/6-31G* molecular orbitals near the HOMO and LUMO orbitals. Orbital number in the HF/6-31G* energy ordering is given, along with D_{2h} symmetry labels.

Figure 5 shows the HF molecular orbitals and energies near the HOMO and LUMO orbitals. These orbitals look visually identical in both the 3-21G and 6-31G* basis sets. Table 6 lists the significant excitations (>0.2 coefficients) that primarily describe each state.

Table 6: Leading coefficients and orbitals in the excited states descriptions at EOM-CCSD/3-21G

Energy (eV)	Osc (f)	Major configurations	
		Transition (all transitions are to the LUMO $7b_{2g}/b_{3g}$ unless specified otherwise)	Coefficients (all > 0.2)
2.42	0.41	$4a_u \longrightarrow$ LUMO $7b_{2g}/b_{3g}$	-0.60
		$9b_{1u}$	0.21
4.34	1.07	$7b_{1u}$	0.40
		$9b_{1u}$	-0.31
		$3a_u$	-0.25
		$4a_u$	-0.24
4.52	0.22	$3a_u$	0.49
		$9b_{1u}$	-0.29
4.57 (B_{1u})	0.01	$30b_{2u}/b_{3u}$	-0.42
		$25b_{1g} \longrightarrow 5a_u$	0.21
4.82	0.40	$9b_{1u}$	-0.33
		$8b_{1u}$	0.29
5.06	0.15	$7b_{1u}$	0.40
		$8b_{1u}$	0.31
		$3a_u$	0.23
5.30	0.10	$4a_u \longrightarrow 8b_{2g}/b_{3g}$	0.60
5.70	0.18	$8b_{1u}$	0.41
		$2a_u$	0.40
6.16	0.11	$2a_u$	0.41
		$4a_u \longrightarrow 9b_{2g}/b_{3g}$	0.27
6.85 (B_{1u})	0.01	$37a_g \longrightarrow 10b_{1u}$	-0.47
6.95	0.05	$6b_{2g}/b_{3g} \longrightarrow 5a_u$	-0.52
7.00	0.68	$4a_u \longrightarrow 8b_{2g}/b_{3g}$	0.52
		$6b_{2g}/b_{3g} \longrightarrow 11b_{1u}$	-0.23

Table 7 benchmarks the effects of freezing orbitals beyond the chemical core on the calculated excitation energies.

Table 7: EOM-CCSD/3-21G energies (in eV), with some additional occupied and/or virtual orbitals frozen, benchmarked against standard calculations where only the chemical core was frozen (shaded).

	EOM-CCSD/3-21G energies (eV)							
Orbitals:								
#active occ	98	98	98	75	50	75	32	18
#active virt	272	167	100	272	272	167	167	37
#frozen occ	49 (core)	49 (core)	49 (core)	72	97	72	115	129
#frozen virt	0	105	172	0	0	105	105	235
				B_{2u}/B_{3u} states				
	2.42	2.49	2.50	2.43	2.46	2.49	2.60	2.55
	4.34	4.31	4.39	4.37	4.46	4.35	4.61	4.63
	4.52	4.53	4.57	4.54	4.60	4.55	4.74	4.96
	4.82	4.85	4.88	4.83	4.88	4.86	4.97	5.08
	5.06	5.09	---	5.07	5.12	5.09	5.20	5.21
				B_{1u} states				
	4.57	4.60	4.61	4.69	5.22	4.72	5.24	5.70

These results show that the energies of the B_{2u}/B_{3u} excited states are weakly dependent on choice of active space, with active space truncation errors averaging only 0.25 eV when 87% of orbitals were frozen in the final column above. Freezing orbitals consistently favours the ground state even further over the excited states, thus increasing the systemic difference in energy between them. The B_{1u} excited state shows a strong dependence on the number of occupied orbitals in the active space, and freezing 97 or more occupied orbitals yields active space truncation errors above 0.6 eV.

EOM-CCSD/6-31G* calculations were made computationally feasible by freezing 312 virtual orbitals in addition to the 49 chemical core orbitals. As Table 8 shows, improving the basis from 3-21G to 6-31G* brought the energies of the excited states closer to the experimental values by around 0.1 eV per state on average.

Table 8: CCSD excitation energies (eV) in different basis sets.

	3-21G	3-21G*	6-31G*	cc-pVDZ
Orbitals:				
#active occ	98	98	98	50
#active virt	272	307	167	107
#frozen occ	49	49	49	97
#frozen virt	0	172	312	475
Experimental	Computational - B_{2u}/B_{3u} states			
1.89	2.42	2.35	2.39	2.76
3.71	4.34	4.30	4.29	4.73
3.74	4.52	4.50	4.45	
4.42	4.82	4.77	4.71	
4.70	5.06	5.01	4.94	
4.88	5.30		5.02	
5.10	5.70		5.58	
	Computational - B_{1u} states			
3.99	4.57		4.86	

The EOM-CCSD/3-21G energies were systemically too high by around 0.6 eV, and increasing the size of the basis set to 6-31G* caused a reduction in calculated excited state energies by 0-0.35 eV for all the B_{2u}/B_{3u} states, mostly ~0.1 eV. It also caused the second excited state to substantially change its character in terms of the excitations of which it is composed, and subsequently the second and third excited states reordered themselves in energy. Experimentally the correct ordering of these two states is difficult to determine, because one is a very intense but broad transition, while the other is a much sharper but less intense peak within the broad band. So it is unclear exactly where the vertical electronic transition of the broad band actually lies. However the effect of the larger basis on the B_{1u} state moved that energy away from the experimental value by 0.3 eV, although this may have been a function of freezing additional orbitals. The small energetic effect (~0.1 eV) resulting from an increase in basis set from 3-21G to 6-31G* is consistent with previous work on related porphyrin system that saw EOM-CC excitation energies change by a similar amount when polarisation functions were included in the basis.[61]

Rydberg states (states predominately described by excitations to Rydberg orbitals) require diffuse functions to be present in the basis set in order to achieve accurate energies for state transitions. Christiansen *et al* found an

effective way to describe Rydberg states without greatly increasing basis set size was to place very diffuse functions on the molecule's centre of mass.[62] On their system this lowered Rydberg state energies by around 2 eV compared to a basis set that lacked diffuse functions. We used the *s*, *p* and *d* augmented functions from the aug-cc-pVTZ basis set for the Zn atom, and added a second set of those functions with halved exponents, to provide very diffuse functions appropriate for a description of Rydberg states. As Table 9 shows, states in the Q, B, N and L bands were altered by at most 0.1 eV in energy, while states in the highest energy, C, band were affected by around 0.4 eV.

Table 9: Effects on excited state energies of adding diffuse functions to the centre of mass (CM). Energy changes > 0.15 eV in bold. The 49 chemical core orbitals were frozen during the calculations as well as some virtual orbitals.

		CCSD/3-21G	CM-augmented CCSD/3-21G	Δ eV:
Orbitals:				
#active occ		98	98	
#active virt		167	99	
#frozen occ		49	49	
#frozen virt		105	191	
Experiment		Computational - B_{2u}/B_{3u} states		
1.89	Q	2.49	2.43	0.07
3.71	B1	4.31	4.37	-0.06
3.74	B2	4.53	4.48	0.05
4.42	N1	4.85	4.80	0.05
4.70	N2	5.09	5.02	0.07
4.88	L1	5.36	5.26	0.10
5.10	L2	5.77	5.70	0.07
5.33	L3	6.20	6.15	0.05
5.92	C2	6.97	6.54	0.42
		Computational - B_{1u} states		
3.99	B3	4.60	4.70	-0.10
5.62	C1	6.23	5.86	0.37

Both the C1 and C2 states were lowered in energy by around 0.4 eV with the addition of diffuse functions placed at the molecule's centre. We were unable to converge the C3 state using this basis. The excitation coefficients of the C2 state reveal that it is partially Rydberg in nature, with the coefficients comprising the transitions to the Rydberg orbital not being the dominant transition (the highest coefficient of excitation to a Rydberg orbital is 0.34, while there is a 0.48 coefficient of excitation to a non-Rydberg virtual orbital). The C1 state, of B_{1u}

symmetry, likewise has a partial Rydberg character, although a low transition intensity.

Table 10 shows the results of different triples-corrected, EOM-CCSD(T)/3-21G, calculations. The different non-iterative triples algorithms implemented in the computational chemistry codes we used gave very different results on this system. Most triples algorithms corrected the energies in the wrong direction – away from the experimental results rather than closer to them – and sometimes introduced errors of multiple electron-volts.

Table 10: EOM-CCSD(T)/3-21G calculations using different algorithms

	CCSD	QChem CCSD(fT)[59]	AcesII CCSDT-1[52]	GAMESS CR-EOML (2,3),D[54]	NWChem CR-EOM-CCSD(T)[51]	GAMESS CR-EOML DEL(IID)[53,60]
Orbitals:						
#active occ	98	98	98	50	98	50
#active virt	272	167	19	100	167	100
#frozen occ	49	49	49	97	49	97
#frozen virt	0	105	253	172	105	172
Experimental	Computational - B_{2u}/B_{3u} states					
1.89	2.42	5.03	3.18	3.16	2.24	2.09
3.71	4.34		5.35	5.12	4.13	3.93
3.74	4.52		5.54	5.27	4.32	4.07
4.42	4.82			5.56	4.65	4.38
4.70	5.06				4.87	4.47
4.88	5.30					4.72
5.10	5.70					5.22
5.33	6.16					5.66
5.92	6.95					6.16
6.00	7.00					6.43
	Computational - B_{1u} states					
3.99	4.57	7.27	6.08		4.49	4.11*
5.62	6.85					5.82*

* Only 72 occupied orbitals were frozen in this calculation, as Table 7 indicated B_{1u} state energies are sensitive to having too many frozen occupied orbitals.

Only the most recently developed algorithms that rigorously balance all types of correlation energy in both the ground and excited state corrected the energy towards the experimental values. The most theoretically rigorous (and most recently developed, 2011) of the corrections, the GAMESS CR-EOML DEL(IID) implementation, corrects most states by around 0.5 eV, bringing them to within

0.26 eV of the experimental values on average, with a 0.4 eV maximum error. Figure 6 shows a plot of these against the experimental spectrum.

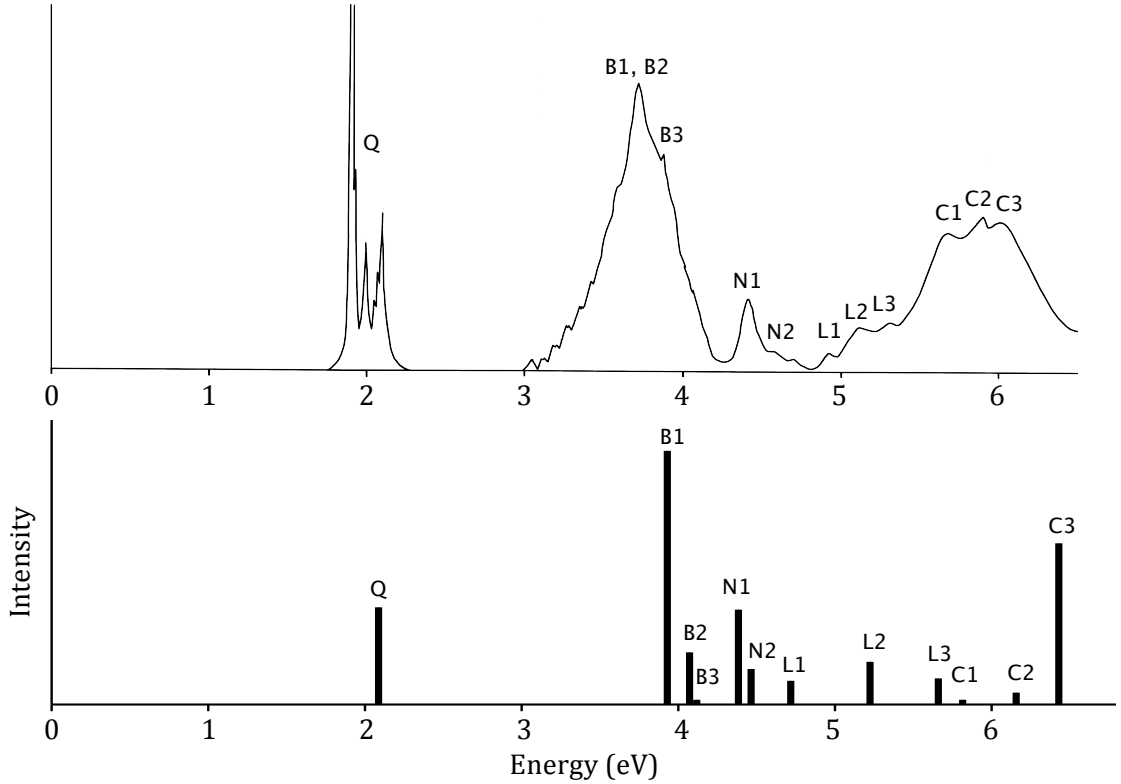


Figure 6: Calculated EOM-CCSD(T)/3-21G spectrum, using the recently developed CR-EOML DEL(IID) corrections implemented in GAMESS, compared to the experimental spectrum.

The energetic accuracy and relative oscillator peaks are enough to assign all people unambiguously, except for L3 and C1. Table 9 showed that the C-band peaks are lowered in energy by up to 0.4 eV at the CCSD level of theory when diffuse functions are included in the basis. That offset is not included in the CCSD(T) spectrum shown in Figure 6 above. Were it to be included, C1 and L3 would be reordered, which would agree better with the experimental spectrum in terms of oscillator strength, as experimentally the C1 transition appears stronger than the L3 transition.

Static and dynamic correlation

To estimate the importance of dynamic correlation in accurately predicting the electronic excitation spectrum of ZnPc, we compare coupled cluster results in a flexible AO basis (CCSD/3-21G) with those obtained in a minimal basis.[63,64] For the minimal basis we used STO-6G but with 6-31G on Zn and with the five highest virtual orbitals frozen. This was necessary as CCSD/MBS calculations with the STO-6G (and STO-3G) basis were not converging, which we traced to these bases' inadequate descriptions of the *3d* metal orbitals, and we were hence able to resolve this by increasing the basis size on the metal ion. Table 11 lists these results, and Figure 6 shows a spectrum of them. Differences in energy and oscillator strength make it difficult to visually match the state transitions shown in Figure 6. However it was possible to match up the states based on their dominant electron configuration and coefficients, producing the assignments listed in Table 11.

Table 11: EOM-CCSD results for a minimal basis set (MBS) versus a 3-21G basis. (States have been matched based on dominant electron configurations and coefficients) The difference between these results approximates the dynamic correlation energy involved in each excited state relative to the ground state.

EOM-CCSD/3-21G		EOM-CCSD/MBS		Difference (eV)
Energy (eV)	*f*	Energy (eV)	*f*	
2.42	0.41	2.77	0.35	0.35
4.34	1.07	4.70	1.19	0.36
4.52	0.22	5.18	0.01	0.66
4.57 (B_{1u})	0.01	4.57	0.01	0.00
4.82	0.40	5.40	0.23	0.58
5.06	0.15	5.85	0.17	0.79
5.30	0.10	6.36	0.01	1.06
5.70	0.18	6.70	0.54	1.00
6.85 (B_{1u})	0.01	6.51	0.11	-0.34
6.95	0.05	6.63	0.58	-0.37

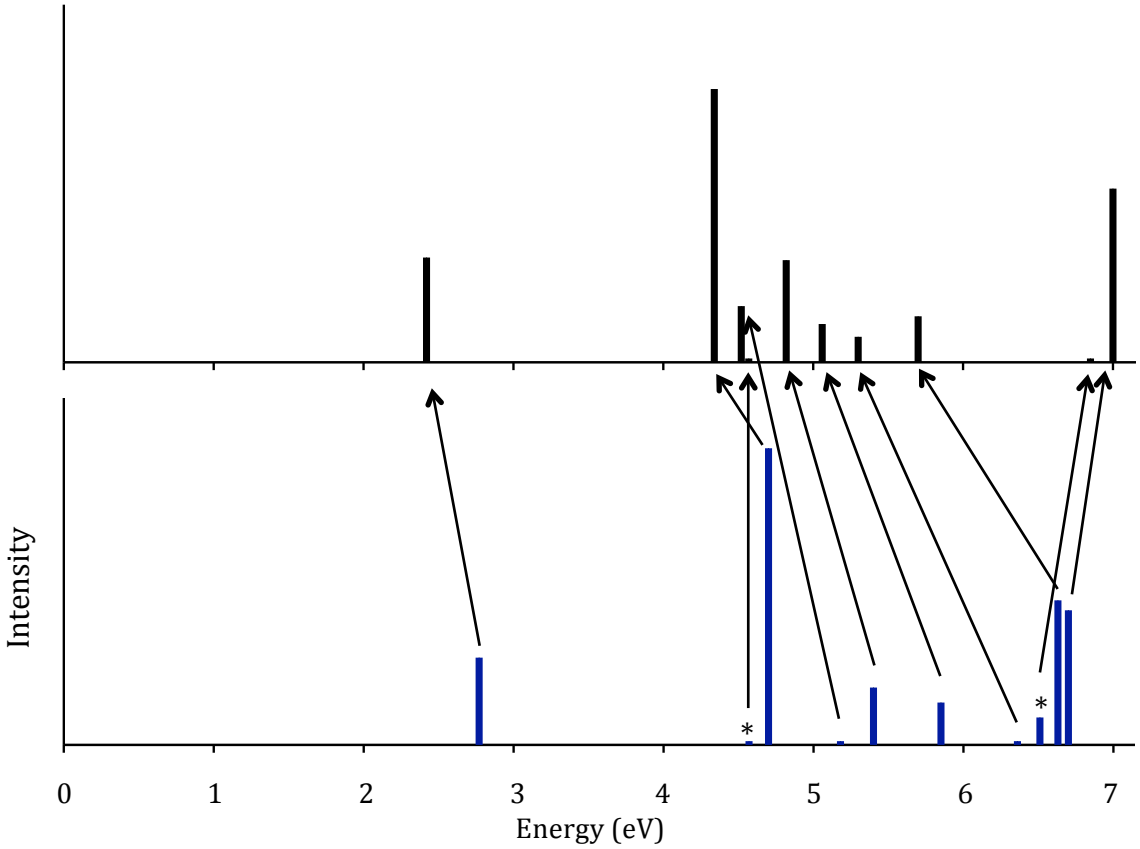


Figure 7: Calculated EOM-CCSD spectrum comparing results for a minimal basis set (blue, bottom) matched with states in the 3-21G basis (black, top). Arrows connect identical states based on comparison of the dominant electron configurations and coefficients. B_{1u} peaks are marked with asterisks.

In Figure 7, the first two high intensity transitions in the minimal basis set calculation can be visibly matched against the 3-21G calculation (the Q transition and the primary transition in the B band) which is confirmed by a comparison of the dominant electron configurations. The other transitions cannot be easily matched visually, due to different transition moments - which result from using an inflexible atomic orbital basis. However when the states are matched using the electron configurations, we see the arrows connecting them in Figure 7 are roughly parallel for the most part, suggesting that dynamic correlation affects the energies of these states in consistent ways. The exceptions to this rule are the two states of different symmetry (B_{1u}), and the very highest energy state, whose partial Rydberg character is likely the reason for the high error in the minimal basis calculations. Thus differences in dynamic correlation energy between states appear to be having little impact on the spectrum.

Static correlation, by contrast, appears to play a major role in this system. With reference to the orbital energies shown earlier in Figure 5, we see multiple orbitals that are energetically close to the HOMO-1. The consequences of this are that when an excitation occurs out of these orbitals, the resulting state can be strongly multireference, as indicated in Table 6. Therefore, the total energy of each excited state includes a significant fraction of static correlation energy. By contrast, the ground state is almost completely single-reference in character due to the HOMO being well-separated from the LUMO. Therefore any method incapable of completely recovering the static correlation energy will prejudice the ground state over the excited states, as it gives a better description of the ground state than of the excited states. We can ascribe the 0.6 offset we applied to our EOM-CCSD results to this effect, as the CCSD method recovers much of the static correlation energy but not all.[63]

Conclusions

The EOM-CCSD/3-21G method accurately predicts the locations of peaks in the Zinc Phthalocyanine spectrum to an average accuracy of 0.22 eV, although with a systemic offset error of 0.6 eV of the excited states relative to the ground state. This is an uncharacteristically high level of error for the EOM-CCSD method, which usually yields errors of < 0.2 eV, albeit on much smaller systems. Increasing the basis set size from 3-21G to 6-31G* reduced the systemic offset error by about 0.1 eV. Including a robust non-iterative triples correction removed the systemic offset error completely and reduced the total error to an average of 0.26 eV per state. Therefore EOM-CCSD/3-21G with a δ_{HD} triples correction is expected to be a promising method for generally predicting electronic excitation spectra of metallophthalocyanines.

References

(1) Löbbert, G. In *Ullmann's Encyclopedia of Industrial Chemistry*; Wiley-VCH: Weinheim, 2002.

(2) Taffa, D. H.; Kathiresan, M.; Arnold, T.; Walder, L.; Erbacher, M.; Bauer, D.; Montforts, F. P.; Nordmann, J.; Haase, M. *J Photoch Photobio A* **2010**, *216*, 35.

(3) Rand, B. P.; Genoe, J.; Heremans, P.; Poortmans, J. *Prog Photovoltaics* **2007**, *15*, 659.

(4) Kushto, G. P.; Makinen, A. J.; Lane, P. A. *Ieee J Sel Top Quant* **2010**, *16*, 1552.

(5) Chawla, P.; Tripathi, M. *Int J Energ Res* **2015**, *39*, 1579.

(6) Armstrong, N. R.; Wang, W. N.; Alloway, D. M.; Placencia, D.; Ratcliff, E.; Brumbach, M. *Macromol Rapid Comm* **2009**, *30*, 717.

(7) Li, L. S.; Guan, M.; Cao, G. H.; Li, Y. Y.; Zeng, Y. P. *Displays* **2012**, *33*, 17.

(8) Van Slyke, S. A.; Chen, C. H.; Tang, C. W. *App. Phys. Lett.* **1996**, *69*, 2160.

(9) Ogbodu, R. O.; Nyokong, T. *Spectrochim Acta A* **2015**, *151*, 174.

(10) Banfi, S.; Caruso, E.; Buccafurni, L.; Ravizza, R.; Gariboldi, M.; Monti, E. *J Organomet Chem* **2007**, *692*, 1269.

(11) Gregory, P. *Journal of Porphyrins and Phthalocyanines* **2000**, *4*, 432.

(12) Vancott, T. C.; Rose, J. L.; Misener, G. C.; Williamson, B. E.; Schrimpf, A. E.; Boyle, M. E.; Schatz, P. N. *J. Phys. Chem.* **1989**, *93*, 2999.

(13) Grobosch, M.; Schmidt, C.; Kraus, R.; Knupfer, M. *Org Electron* **2010**, *11*, 1483.

(14) Schechtman B. H.; Spicer, W. E. *J Mol Spectrosc* **1970**, *33*, 28.

(15) Seoudi, R.; El-Bahy, G. S.; El Sayed, Z. A. *Opt Mater* **2006**, *29*, 304.

(16) Edwards, L.; Dolphin, D. H.; Gouterman, M. *J Mol Spectrosc* **1970**, *35*, 90.

(17) Edwards, L.; Gouterman, M. *J Mol Spectrosc* **1970**, *33*, 292.

(18) Edwards, L.; Gouterman, M.; Rose, C. B. *J Am Chem Soc* **1976**, *98*, 7638.

(19) Gurol, I.; Durmus, M.; Ahsena, V.; Nyokong, T. *Dalton T* **2007**, 3782.

(20) Mack, J.; Stillman, M. J. *J. Phys. Chem.* **1995**, *99*, 7935.

(21) Metcalf, D. H.; Vancott, T. C.; Snyder, S. W.; Schatz, P. N.; Williamson, B. E. *J. Phys. Chem.* **1990**, *94*, 2828.

(22) Nyokong, T.; Gasyna, Z.; Stillman, M. J. *Inorg Chem* **1987**, *26*, 1087.

(23) Gouterman, M. In *Porphrins*; D., Dophin, Ed.; Academic Press: New York, 1978; Vol. 3, p 1.

(24) Gouterman, M.; Snyder, L. C.; Wagniere, G. H. *J Mol Spectrosc* **1963**, *11*, 108.

(25) Marom, N.; Kronik, L. *Applied Physics A - Materials Science & Processing* **2009**, *95*, 165.

(26) Grobosch, M.; Mahns, B.; Loose, C.; Friedrich, R.; Schmidt, C.; Kortus, J.; Knupfer, M. *Chem. Phys. Lett.* **2011**, *505*, 122.

(27) Friedrich, R.; Hahn, T.; Kortus, J.; Fronk, M.; Haidu, F.; Salvan, G.; Zahn, D. R. T.; Schlesinger, M.; Mehring, M.; Roth, F.; Mahns, B.; Knupfer, M. *J. Chem. Phys.* **2012**, *136*.

(28) Ueno, L. T.; Jayme, C. C.; Silva, L. R.; Pereira, E. B.; de Oliveira, S. M.; Machado, A. E. H. *J Brazil Chem Soc* **2012**, *23*, 2237.

(29) Mack, J.; Stillman, M. J. *J Am Chem Soc* **1994**, *116*, 1292.

(30) Mack, J.; Stillman, M. J. *Inorg Chem* **1997**, *36*, 413.

(31) Mack, J.; Stillman, M. J. *Coordin Chem Rev* **2001**, *219*, 993.

(32) Mack, J.; Stillman, M. J. *Journal of Porphyrins and Phthalocyanines* **2001**, *5*, 67.

(33) Ridley, J.; Zerner, M. *Theor Chim Acta* **1973**, *32*, 111.

(34) Pople, J. A.; Beveridg.Dl; Dobosh, P. A. *J. Chem. Phys.* **1967**, *47*, 2026.

(35) Liao, M. S.; Scheiner, S. *J. Chem. Phys.* **2001**, *114*, 9780.

(36) Liao, M. S.; Watts, J. D.; Huang, M. J.; Gorun, S. M.; Kar, T.; Scheiner, S. *J Chem Theory Comput* **2005**, *1*, 1201.

(37) Mack, J.; Kobayashi, N.; Stillman, M. J. *Journal of Porphyrins and Phthalocyanines* **2006**, *10*, 1219.

(38) Nguyen, K. A.; Pachter, R. *J. Chem. Phys.* **2001**, *114*, 10757.

(39) Theisen, R. F.; Huang, L.; Fleetham, T.; Adams, J. B.; Li, J. *J. Chem. Phys.* **2015**, *142*.

(40) Ricciardi, G.; Rosa, A.; Baerends, E. J. *J. Phys. Chem. A* **2001**, *105*, 5242.

(41) Chaudhuri, R. K.; Freed, K. F.; Chattopadhyay, S.; Mahapatra, U. S. *J. Chem. Phys.* **2011**, *135*, 084118.

(42) Bartlett, R. J. In *Modern Electronic Structure Theory*; 2nd ed.; Yarkony, D.R., Ed.; World Scientific: 1995, p 1047.

(43) Hirata, S.; Nooijen, M.; Bartlett, R. J. *Chem. Phys. Lett.* **2000**, *326*, 255.

(44) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(45) Gordon, M. S.; Schmidt, M.W. In *Theory and Applications of Computational Chemistry*; Dykstra, C.E.; Frenking, G.; Kim, K.S.; Scuseria, G.E., Ed.; Elsevier: Amsterdam, 2005.

(46) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. *Comput Phys Commun* **2010**, *181*, 1477.

(47) Stanton, J. F.; Gauss, J.; Perera, S. A.; Watts, J. D.; Yau, A. D.; Nooijen, M.; Oliphant, N.; Szalay, P. G.; Lauderdale, W. J.; Gwaltney, S. R.; Beck, S.; Balkova, A.; Bernholdt, D. E.; Baeck, K. K.; Rozyczko, P.; Sekino, H.; Huber, C.; Pittner, J.; Cencek, W.; Taylor, D.; Bartlett, R. J. *ACES II*; University of Florida: Florida.

(48) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C. P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys Chem Chem Phys* **2006**, *8*, 3172.

(49) Bozkaya, U.; Schaefer, H. F. *J. Chem. Phys.* **2012**, *136*.

(50) Kowalski, K. *Chem. Phys. Lett.* **2005**, *411*, 306.

(51) Kowalski, K.; Piecuch, P. *J. Chem. Phys.* **2004**, *120*, 1715.

(52) Lee, Y. S.; Kucharski, S. A.; Bartlett, R. J. *J. Chem. Phys.* **1984**, *81*, 5906.

(53) Lutz, J. J. *Development and Applications of Coupled-Cluster Methods and Potential Energy Surface Extrapolation Schemes*, PhD Thesis, Michigan State University, 2011.

(54) Piecuch, P.; Wloch, M. *J. Chem. Phys.* **2005**, *123*.

(55) Sauer, S. P. A.; Schreiber, M.; Silva, M. R.; Thiel, W. *J Chem Theory Comput* **2009**, *5*, 555.

(56) Watts, J. D.; Bartlett, R. J. *J. Chem. Phys.* **1994**, *101*, 3073.

(57) Watts, J. D.; Bartlett, R. J. *Chem. Phys. Lett.* **1995**, *233*, 81.

(58) Watts, J. D.; Bartlett, R. J. *Chem. Phys. Lett.* **1996**, *258*, 581.

(59) Hirata, S.; Fan, P. D.; Auer, A. A.; Nooijen, M.; Piecuch, P. *J. Chem. Phys.* **2004**, *121*, 12197.

(60) Piecuch, P.; Gour, J. R.; Wloch, M. *Int J Quantum Chem* **2009**, *109*, 3268.

(61) Serrano-Andres, L.; Merchan, M.; Rubio, M.; Roos, B. O. *Chem. Phys. Lett.* **1998**, *295*, 195.

(62) Christiansen, O.; Koch, H.; Jorgensen, P.; Helgaker, T. *Chem. Phys. Lett.* **1996**, *263*, 530.

(63) Crittenden, D. L. *J. Phys. Chem. A* **2013**, *117*, 3852.

(64) Wallace, A. J.; Crittenden, D. L. *J. Phys. Chem. A* **2014**, *118*, 2138.

5. The Performance of CASSCF and CASCI Based Methods for Modeling the Electronic Excitation Spectrum of ZnPc

Abstract

In addition to the EOM-CCSD based methods discussed in the previous chapter, we have explored a number of other multireference methods to calculate the electronic excitation spectrum of Zinc Phthalocyanine. These results were unsatisfactory; CASSCF and CASCI methods produced results that were to 1-2 eV in error compared to experimental data, and the generated states were out of order compared to the EOM-CCSD results. MRMP2 excitation energies were numerically unstable with respect to the number of states included, and with respect to the energy denominator offset that was introduced to address the intruder state problem.

Introduction

Complete active space self-consistent field (CASSCF) based methods[1,2] are often used for modeling the excited states of multireference systems.[3-6] They have two potential advantages over the EOM-CC family of methods. The first is that calculations can usually be run using larger basis sets, as the methods scale primarily with regard to the size of the active space rather than with the size of the basis set. Second is that a perturbation-theory energy correction can be applied to improve the accuracy of the excitation energies. These methods are referred to as multireference Møller-Plesset perturbation theory (MRMP), or complete active space perturbation theory (CASPT), or multiconfiguration quasi-degenerate perturbation theory (MCQDPT). The perturbation theory expansion is nearly always truncated at second order, giving MRMP2, CASPT2, and MCQDPT2 respectively. The disadvantages of these methods are: A) a limited number of orbitals can be included in the active space; B) the choice of which orbitals to include can be arbitrary and affects the calculated energies; C) the perturbation step can suffer from numerical instabilities due to 'intruder states',

if the included orbitals do not provide a sufficiently accurate first-order description of the relevant states. The 'intruder state' problem is a well-understood problem that these methods suffer from on complex systems,[6-8] and arises from near-zero terms in the denominator of the perturbation theory energy correction expression. These result in large and inaccurate energy corrections, and can lead to incorrect state reorderings when high-energy states that are badly-described by the chosen active space receive unduly large perturbation theory energy corrections and thus appear low in energy in the final output, 'intruding' into the midst of the other lower-energy states. Various workarounds for the problem exist, the most common of which is to add a small arbitrary term to the energy denominator to stop the division by near-zero occurring and therefore limit the maximum size of each perturbation theory correction term.[7,8]

Methods

All CASSCF,[2,9] CASCI, MRMP2[10-14], and MCQDPT2[13,14] calculations in this section were performed with the GAMESS package.[15,16] Molecular orbitals were visualized using MacMolPlt.[17] The geometry used for the metallophthalocyanines is given in Appendix 3. Absolute energies and details of the orbitals included in active spaces are given in Appendix 4.

For any CASSCF-based calculation, it is necessary to carefully select certain orbitals to be part of the active space. These calculations are not 'black box' in the same way EOM-CCSD type calculations are, and require careful choice on the part of the user. These calculations also scale factorially with increased size of active space, which limits how many orbitals it is practical to include in the active space, with 12 orbitals and 12 electrons usually being close to the limit. Our EOM-CCSD calculations in the previous chapter identified the primary transitions that contribute to the excited states. Those calculations found that the first 12 symmetry-allowed excited state transitions were dominated by transitions from the 10 highest energy occupied orbitals and to 8 of the virtual orbitals (the first 4 lowest virtuals in energy, plus the next two b_{2g}/b_{3g} pairs). Those orbitals are depicted in Figure 1. Past TD-DFT calculations from the

literature also agree that this set, or a subset thereof, are the dominant orbital transitions for the excited states in the spectrum of this system.[18,19]

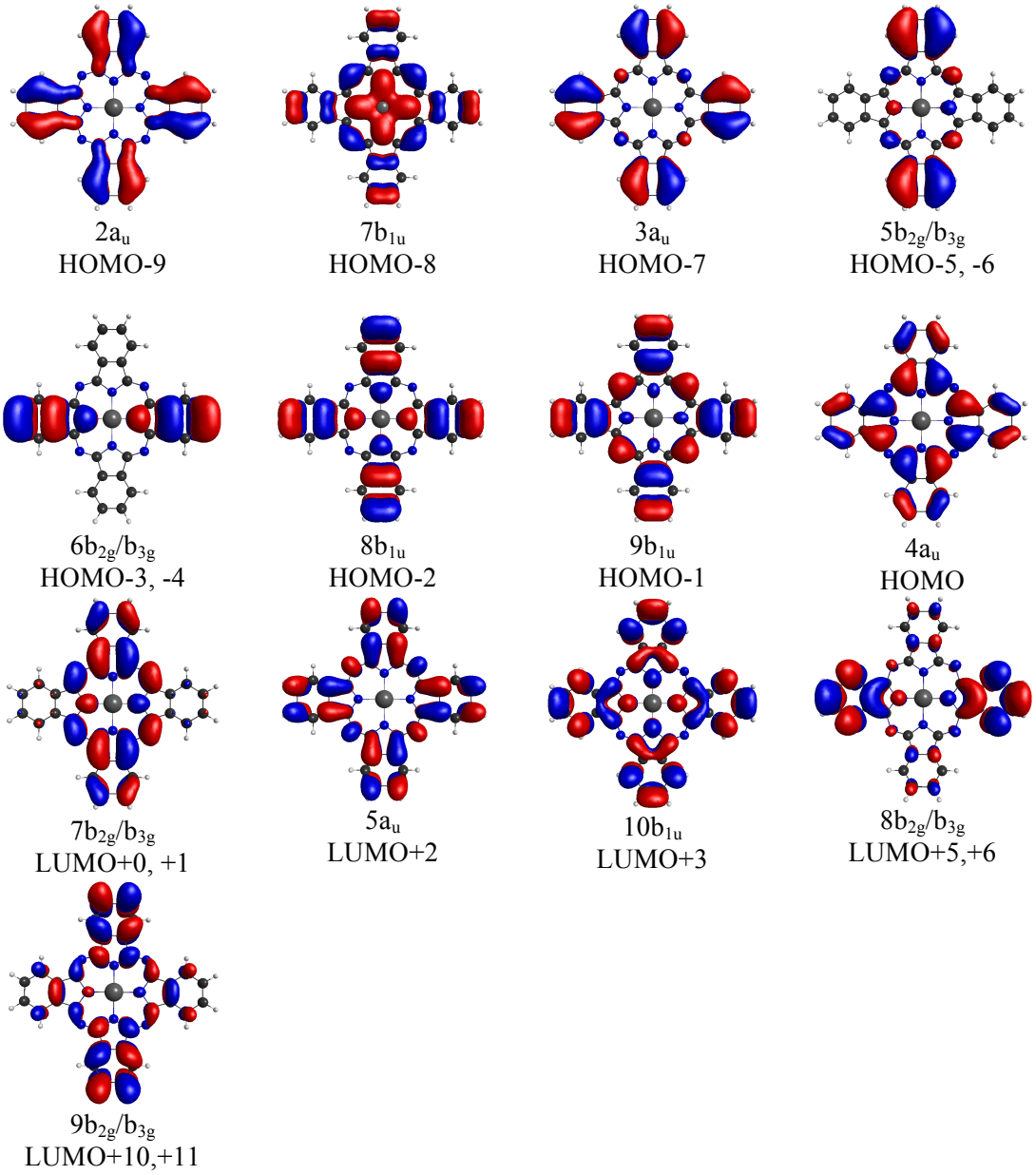


Figure 1: ZnPc molecular orbitals that dominate the orbital transitions of the first 12 excited states. D_{2h} symmetry labels are given, in the HF/6-31G* energy ordering.

Unfortunately, running a complete active space calculation with all 18 of the orbitals included is computationally infeasible, and so it is necessary to exclude some orbitals, and therefore states, from the calculation. The fact that the B_{2u}/B_{3u} states are degenerate also presents a practical problem for CASSCF calculations in GAMESS since GAMESS' CASSCF algorithm will optimize states of a single

symmetry but not of two symmetries simultaneously – and optimizing the orbitals for B_{2u} states alone would lead to symmetry breaking. Therefore symmetry often has to be turned off entirely during these calculations, yielding states of all symmetries and losing improvements in speed and memory usage that symmetry-specific algorithms offer.

CASSCF basis set convergence

To quantify the effect of different basis sets with the CASSCF and MRMP2 methods, we can use the Q transition as a test case. Table 1 tabulates the energies of the Q transition in CASSCF and MRMP2 calculations performed using different atomic orbital basis sets.

Table 1: CASSCF and MRMP2 results for the energies of the Q transition in different bases (energies in eVs). Experimental transition value is 1.89eV in an argon matrix. Active space consists of 5 orbitals (HOMO, HOMO-1, doubly degenerate LUMO, and LUMO+2).

Method:	CASSCF	CASSCF	CASSCF	CASSCF	MRMP2	MRMP2	MRMP2
Basis set:	3-21G	3-21G*	cc-pVDZ	cc-pVTZ	3-21G	3-21G*	cc-pVDZ
Transition energy (eV):	3.11	3.31	2.99	3.05	2.00	1.90	1.77

CASSCF over-estimates the transition energy by more than 1 eV, irrespective of basis set completeness. This problem was also observed by Chaudhuri *et al.* with regard to similar calculations on the related free base porphyrin molecule.[20] The effect of different basis sets on the energy of the Q transition is about 0.3 eV. At the MRMP2 level, all basis sets yield transition energies within 0.12 eV of the experimental value.

Results and Discussion

CASSCF calculations

The first 6 B2u/B3u excited states identified in the EOM-CCSD calculation can be described using excitations from only 4 occupied and 7 virtual orbitals, using orbitals $7b_{1u}$, $3a_u$, $9b_{1u}$ and $4a_u$ of the occupied orbitals depicted in Figure 1, and omitting $10b_{1u}$ from the virtuals. Table 2 shows the results of a CASSCF calculation that includes those orbitals in the active space and also included the

highest energy metal d-orbital ($d_{x^2-y^2}$), to double-check for involvement from the metal, but no excitations from that orbital were observed.

Table 2: CASSCF/cc-pVDZ calculation results using an active space of 12 orbitals and 10 electrons. The generated states are incorrectly ordered compared to the EOM-CCSD calculations in the previous chapter. Symmetry-allowed transitions in bold.

Experimental energies (eVs):	Calculated energies (eVs):	Excited state symmetry	Corresponding excited state number in EOM-CCSD (Symmetry allowed transitions only)
1.89	3.02	**B_{2u}/B_{3u}**	#1
	3.54	A_g (double excitation)	
	4.58	B_{1g} (double excitation)	
	4.78	A_g (double excitation)	
3.71	5.74	**B_{2u}/B_{3u}**	#6
	6.03	A_g	
3.74	6.13	**B_{1u}**	#4
4.42	6.31	**B_{2u}/B_{3u}**	#2
	6.40	B_{2g}/B_{3g}	
4.70	6.68	**B_{2u}/B_{3u}**	#3

All predicted energies in this calculation are 1-2 eV higher than experimental values.

The states are out of energy order compared to the EOM-CCSD results reported in the previous chapter. By looking at the orbital transitions comprising the CASSCF and EOM-CCSD excited states, it is possible to match the states against each other. EOM-CCSD excited state #6 (predominantly a transition from the HOMO to the LUMO+5,+6 orbitals) is the second-lowest energy of the allowed transitions in the CASSCF calculation. State #4 has also moved out of the correct energy ordering. These results show that the limited size CASSCF active space is providing a poor first-order definition of the states, and that a large perturbation theory energy correction will be necessary to bring these state energies and orderings into agreement with experimental values. However perturbation theory works best when the perturbation is small. Not only will the perturbation theory correction need to correct the energies of the states by up to 2 eV, it will also need to reorder the states, which can cause practical difficulties with intruder states. Thus we anticipate that applying a perturbation theory correction to the results on this system may prove problematic for any symmetry-allowed excited states beyond the first, due to the presence of

symmetry-allowed excited state number 6 being second in the CASSCF energy ordering.

We explored the inclusion of other orbitals in the CASSCF active spaces, but were not able to obtain results that were in better agreement with experimental energies or the EOM-CCSD state orderings than the results given in Table 2 (data not shown).

The CASSCF calculation found a double-excitation to the symmetry-forbidden A_g excited state to be the second excited state. Experimentalists have identified an area of the spectrum in the region around 2.08eV, where the second vibronic band of the first transition (Q) takes an unexpected shape, and they have suggested that there could be a weak electronic transition (which they label Q') in this region of the spectrum that is interacting with the vibronic excitation.[21,22] However TD-DFT studies have failed to provide any plausible explanation for this possible transition,[19] and our EOM-CCSD study identified no electronic transitions in that region of the spectrum. It is conceivable that a symmetry-forbidden transition to this A_g state could explain the shape of the Q' transition. To further investigate this hypothesis, we performed coupled cluster and MRMP2 calculations

EOM-CCSD/3-21G places this A_g transition at 4.23eV, and EOM-CCSD(T)/3-21G at 4.0eV using GAMESS' ∂^{IID} algorithm that was identified in the previous chapter as the most accurate EOM-CCSD(T) algorithm explored. Running an MRMP2 calculation on two states with the active space described in Table 2, puts the energies at 1.80eV for the Q band (1.89eV experimentally) and 2.85eV for the A_g double-excitation, so it does not appear that the doubly-excited A_g state is an appropriate assignment for Q' at 2.08eV.

Complete active space configuration interaction (CASCI)

Due to the similarity between the Hartree-Fock (HF) orbitals and CASSCF orbitals, we explored the possibility of using CASCI calculations[2] that did not

optimize the orbitals. The CASCI calculations allowed us to use larger active spaces and enable symmetry.

Table 3: Results from CASCI calculations using different active spaces and basis sets. The active orbitals are those closest in energy to the LUMO and the HOMO.

Active orbitals						
Occupied:	18	18	38	31	9	11
Virtual:	3	3	3	3	6	6
Basis:	3-21G	cc-pVDZ	3-21G	cc-pVDZ	cc-pVDZ	3-21G
Experimental:	Calculated energies:					
1.89	2.97	2.86	2.91	2.61	2.88	3.02
3.71	5.12	5.05	5.04	4.75	4.97	5.11
3.74	5.51	5.55	5.41	5.26	5.36	5.54
4.42	5.65	5.62	5.55	5.28	5.48	5.66
4.70	6.04	6.08	5.97	5.81	5.83	---

Table 3 shows the results of CASCI calculations performed using different basis sets and active spaces. The first two calculated energy columns are directly comparable - they use the same orbitals but different basis sets, and their energies differ by 0.1 eV at most. This is consistent with our previous findings that CASSCF excitation energies are largely insensitive to basis set. However, the total energies are all too high by 1-1.8 eV compared to experiment, even with the largest active spaces.

The EOM-CCSD calculations in the previous chapter indicated that the five lowest energy excited states are heavily dominated by configurations that have excitations into only the first three virtual orbitals. Including only 3 virtual orbitals in our active space allowed us to include up to 38 occupied orbitals, as the results in Table 3 indicate. Comparing the values in the three final columns with the experimental energies show that the calculated energies become less accurate as more virtual orbitals and fewer occupied orbitals are included in the active space for these states. An addition advantage of limiting the number of virtual orbitals is that the 6th excited state can't appear out of place in the energy ordering because the orbitals transitions that describe it are not present in the active space.

It was our experience with additional exploratory CASCI calculations, not reported here, that CASCI results neither agreed with experimental energies nor EOM-CCSD state orderings regardless of choice of active space. This is in concordance with the CASSCF results shown in Table 2.

Multireference Møller-Plesset second-order perturbation theory (MRMP2)

Multireference Møller-Plesset second-order perturbation theory methods (MRMP2[23], MCQDPT2[13,14], and CASPT2[24,25]) are among the most advanced *ab initio* computational chemistry methods in terms of their ability to recover both static and dynamic correlation energies. Using a CASSCF or CASCI starting point, these methods apply a perturbation theory energy correction to the state energies.[3,4,6] We have consistently found that CASCI and CASSCF calculations yield energies that are too high by up to 2 eV on the ZnPc system.

Table 4 gives the calculated energies of different states after a perturbation theory energy correction has been applied. These results indicate numerical instability with regard to the value of the energy-denominator offset (ε) and show divergent MRMP2 and MCQDPT2 results. It was our experience in additional calculations not reported here that the calculated energies of higher energy excited states showed the same instabilities.

Table 4: MRMP2 and MCQDPT2 results with active space comprising 7 active orbitals - the first three virtuals, the top two occupied orbitals, and occupied orbitals $7b_{1u}$ and $1a_u$. The energy-denominator offset (ε) limits the effect of near-zero denominators.

Excited state #		First	Second
Experimental		1.89	3.71
CASSCF		3.0	6.2
MRMP2	ε = 0.0	1.72	2.87
MCQDPT2	ε = 0.0	0.67	3.69
	ε = 0.002	0.84	3.76
	ε = 0.02	1.08	4.03
	ε = 0.05	1.27	4.27

A variation on CASSCF and CASCI methods is the Occupation-Restricted Multiple Active Spaces (ORMAS) method,[26-28] which gives the user fine-grained control over what excitations are allowed among the active orbitals. This theoretically

allows for larger active spaces to be used, as the excitation level can be reduced, in an approach approximately equivalent to a configuration-interaction singles and doubles calculation (CISD). Including a perturbation theory correction gives a model that approximates CISD(T). However, this really corresponds to an incomplete version of EOM-CCSD(T), and it was our experience with exploratory calculations not reported here that the computational requirements of this method were high and that the calculations suffered from the same sorts of issues as the CASSCF and CASCI methods covered in this chapter.

Conclusions

CASSCF and CASCI methods overestimate the existed state energies for this system by over 1 eV compared to experimental results, and generate states out of order with respect to the EOM-CCSD orderings for excited states beyond the first, with the sixth exited state taking the place of the second excited state in the energy ordering. MRMP2 yielded an energy for the first excited state accurate to within 0.12 eV regardless of basis. However, both MRMP2 and MCQDPT2 exhibited errors of up to 1 eV when a second excited state was also calculated. MCQDPT2 energies showed instability with respect the value of the energy-denominator offset (ε), designed to prevent near-zero terms in the denominator and solve the intruder state problem. We believe these issues were occurring as a result of the limited size of the active spaces we were able to use. Achieving experimentally accurate and numerically stable results with these methods requires more orbitals in the active space than is computationally feasible. These active space multireference methods appear to be of little practical value for calculating the electronic excitation spectrum of this molecule. This explains why there have been no publications using those methods to calculate the electronic excitation spectrum of this molecule, despite multiple calls for such work to be performed.[19,22,29]

References

(1) Roos, B. O. In *Adv Chem Phys*; Lawley, K. P., Ed.; Wiley Interscience: New York, 1987; Vol. 69, p 339.

(2) Schmidt, M. W.; Gordon, M. S. *Annu Rev Phys Chem* **1998**, *49*, 233.

(3) Abrams, M. L.; Sherrill, C. D. *J. Phys. Chem. A* **2003**, *107*, 5611.

(4) Azizi, Z.; Roos, B. O.; Veryazov, V. *Phys Chem Chem Phys* **2006**, *8*, 2727.

(5) Olsen, J. *Int J Quantum Chem* **2011**, *111*, 3267.

(6) Pulay, P. *Int J Quantum Chem* **2011**, *111*, 3273.

(7) Camacho, C.; Witek, H. A.; Yamamoto, S. *J. Comput. Chem.* **2009**, *30*, 468.

(8) Witek, H. A.; Choe, Y. K.; Finley, J. P.; Hirao, K. *J. Comput. Chem.* **2002**, *23*, 957.

(9) Szalay, P. G.; Muller, T.; Gidofalvi, G.; Lischka, H.; Shepard, R. *Chem. Rev.* **2012**, *112*, 108.

(10) Hirao, K. *Chem. Phys. Lett.* **1992**, *196*, 397.

(11) Hirao, K. *Chem. Phys. Lett.* **1992**, *190*, 374.

(12) Hirao, K. *Int J Quantum Chem* **1992**, 517.

(13) Nakano, H. *J. Chem. Phys.* **1993**, *99*, 7983.

(14) Nakano, H. *Chem. Phys. Lett.* **1993**, *207*, 372.

(15) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(16) Gordon, M. S.; Schmidt, M.W. In *Theory and Applications of Computational Chemistry*; Dykstra, C.E.; Frenking, G.; Kim, K.S.; Scuseria, G.E., Ed.; Elsevier: Amsterdam, 2005.

(17) Bode, B. M.; Gordon, M. S. *J Mol Graph Model* **1998**, *16*, 133.

(18) Ricciardi, G.; Rosa, A.; Baerends, E. J. *J. Phys. Chem. A* **2001**, *105*, 5242.

(19) Ueno, L. T.; Jayme, C. C.; Silva, L. R.; Pereira, E. B.; de Oliveira, S. M.; Machado, A. E. H. *J Brazil Chem Soc* **2012**, *23*, 2237.

(20) Chaudhuri, R. K.; Freed, K. F.; Chattopadhyay, S.; Mahapatra, U. S. *J. Chem. Phys.* **2011**, *135*, 084118.

(21) Mack, J.; Stillman, M. J. *Coordin Chem Rev* **2001**, *219*, 993.

(22) Vancott, T. C.; Rose, J. L.; Misener, G. C.; Williamson, B. E.; Schrimpf, A. E.; Boyle, M. E.; Schatz, P. N. *J. Phys. Chem.* **1989**, *93*, 2999.

(23) Choe, Y. K.; Nakao, Y.; Hirao, K. *J. Chem. Phys.* **2001**, *115*, 621.

(24) Andersson, K.; Malmqvist, P. A.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218.

(25) Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483.

(26) Ivanic, J. *J. Chem. Phys.* **2003**, *119*, 9377.

(27) Ivanic, J. *J. Chem. Phys.* **2003**, *119*, 9364.

(28) Ivanic, J. *Abstr. Pap. Am. Chem. S.* **2003**, *226*, U283.

(29) Marom, N.; Kronik, L. *Applied Physics A - Materials Science & Processing* **2009**, *95*, 165.

6. Establishing the Gas-Phase Ground States of the Metallophthalocyanines Using Second-Order Multireference Perturbation Theory

Abstract

We explore the nature of the ground states, and near-to-ground states, for the metallophthalocyanines Mn, Fe, and Co using multireference second-order perturbation theory (MRMP2). We determine the ground states to be $^2A_{1g}$ for CoPc with a *3d*-metal orbital occupancy of $(d_{xy})^2(d_{xz},d_{yz})^4(d_{z^2})^1(d_{x^2-y^2})^0$, ^3E_g for FePc with orbital occupancy of $(d_{xy})^2(d_{xz},d_{yz})^3(d_{z^2})^1(d_{x^2-y^2})^0$, and $^4A_{2g}$ for MnPc with orbital occupancy of $(d_{xy})^2(d_{xz},d_{yz})^2(d_{z^2})^1(d_{x^2-y^2})^0$.

Introduction

Previous DFT and spectroscopic studies of iron and manganese phthalocyanine provide conflicting evidence on the orbital occupancy of their ground state electronic configurations.[1-10] This is due to the fact that the central metal ions have near-degenerate *d*-orbitals, so although the orbital occupation of the ligand remains the same as for ZnPc, there are a number of different ways the metal *d*-orbitals can be filled. Several authors have called for the application of multireference methods to these systems in order to settle the problem more convincingly.[8,11,12]

In our previous chapter, we were unable to use a sufficiently large active space in calculations of the electronic spectrum of ZnPc to describe the full electronic absorption spectrum, involving ligand-ligand transitions that take place among 18 π and π* ligand orbitals. However, to find the ground state we need only consider all possible ways of rearranging electrons amongst the five *3d*-metal orbitals and the ligand orbitals closest to the HOMO-LUMO gap, so it is possible to include all relevant orbitals within the active spaces. We therefore anticipate that MRMP2 will accurately model the energies of the ground and near-to-ground states of these systems.

We will compare our results to the existing spectroscopic and DFT literature results. Where appropriate, the influence of environmental effects will also be considered to rationalize differences between gas-phase computational predictions and solid-state experimental observations.

Methods

All CASSCF[13,14] and MRMP2[15-19] calculations in this chapter were performed using a cc-pVDZ basis[20-22] with the GAMESS package.[23,24] MacMolPlt was used to visualize orbitals.[25] The geometry used for the metallophthalocyanines is given in Appendix 3, and absolute energies for all calculations are reported in Appendix 5.

The active spaces for the CASSCF calculations in this chapter comprise the ligand orbitals nearest the HOMO-LUMO, and orbitals of predominantly (>90%) *3d* metal orbital character, as specified in the results section. The unoccupied $d_{x^2-y^2}$ orbital is omitted from most calculations on Mn, Fe and Co phthalocyanines. Some calculations also omit the d_{xy} metal orbital where it remains doubly-occupied in all near-to-ground states. In some calculations, additional virtual orbitals with *3d* metal orbital character were also included in the active space, following the recommendation of Pulay,[26] which were visual duplicates of the included occupied metal orbitals. The composition of the active space for individual calculations is reported in the results section.

All symmetry labels in this chapter are given in D_{4h} symmetry for ease of comparison to published results.

Results

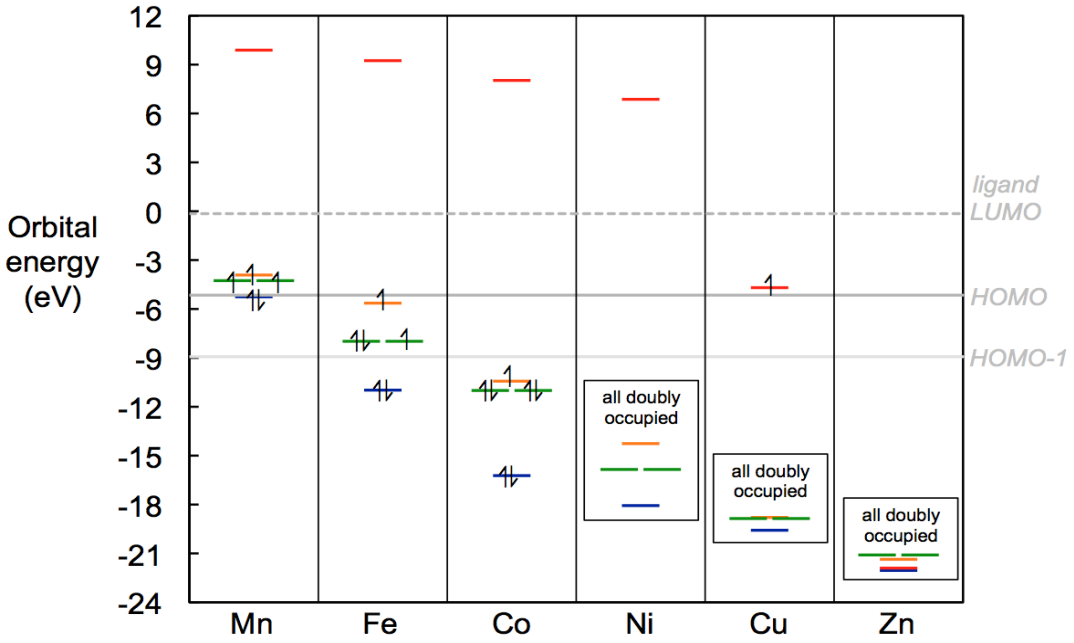


Figure 1: Energies of the *3d* metal orbitals at CASSCF/cc-pVDZ and their occupations in the calculated MRMP2 ground states, with CASSCF active space comprising the five *3d*-metal orbitals. d_{xy} (blue), d_{xz}/d_{yz} (green), d_{z^2} (orange), $d_{x^2-y^2}$ (red).

Figure 1 shows the energies and occupations of the *3d*-metal orbitals for the calculated ground states for each metallophthalocyanine. The ligand HOMO and LUMO are well-separated energetically from one another and from the other ligand orbitals.

A number of experimental studies use ligand field theory to interpret their spectra,[27-31] and assume canonical square planar crystal field energy orderings for the *3d* metal orbitals:

$d_{xz}/d_{yz} \approx d_{z^2} < d_{xy} << d_{x^2-y^2}$

These orbitals energy orderings differ the energy orderings we observed in our CASSCF calculations reported in Figure 1. This difference is presumably a result of the phthalocyanine being a square planar core contained within a macrocyclic system, rather than being a perfectly square planar crystal field of point charges.

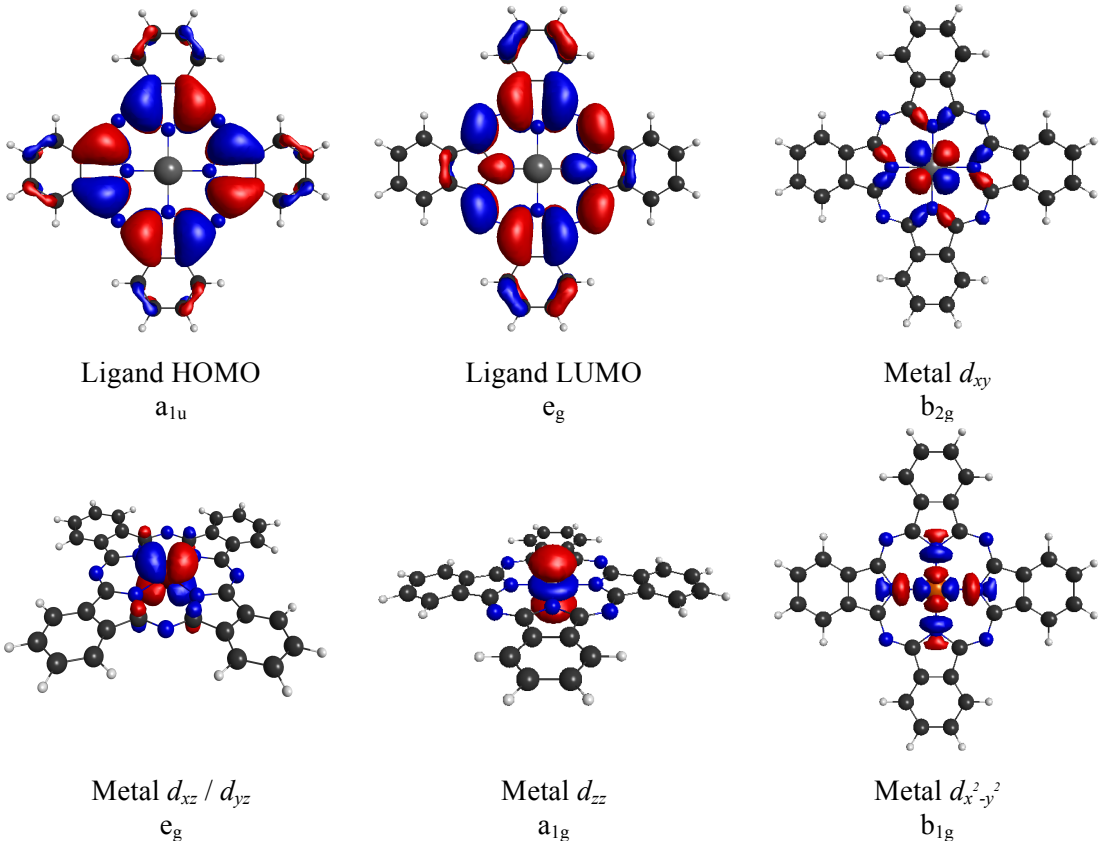


Figure 2: Metal *3d* orbitals, and ligand HOMO and LUMO orbitals. From FePc CASSCF/cc-pVDZ calculations.

Figure 2 shows the ligand HOMO and LUMO orbitals, as well as the *3d*-metal orbitals in the metallophthalocyanines. The orbitals with *3d*-metal character can be readily identified visually as the 3d-metal atomic orbitals have not mixed substantially with the ligand orbitals during the formation of the molecular orbitals.

Ground states of ZnPc, CuPc and NiPc

From Figure 1 we can observe that NiPc and ZnPc have the same occupied orbitals in the vicinity of the HOMO-LUMO gap and that the only difference between these systems will be whether the $d_{x^2-y^2}$ orbital is occupied. In the case of ZnPc, all *3d*-metal orbitals are fully occupied, while in NiPc the $d_{x^2-y^2}$ orbital is unoccupied. These ground state configurations are unanimously agreed upon in the literature.[11,12,32-45]

In CuPc the $d_{x^2-y^2}$ orbital sits between the HOMO and the LUMO, and thus will be singly occupied in the ground state of the system, a configuration that is agreed-upon in the literature.[32,46-53]

Ground state of CoPc

Table 1 gives the CASSCF and MRMP2 energies for the lowest energy states of CoPc. For CoPc, the $d_{x^2-y^2}$ orbital is unoccupied, and the 7 electrons on the metal occupy the four other *3d* orbitals, one of which must therefore be singly occupied. Both CASSCF and MRMP2 models agree that the ground state is $^2A_{1g}$, with the d_{z^2} orbital which is singly occupied. Degenerate ^2E_g excited states have the d_{z^2} orbital fully occupied and the d_{xz} / d_{yz} degenerate pair are one electron short of filled.

Most previous experimental and computational studies concur with this assignment,[54-56] with only a single DFT study predicting that the ^2E_g state is lowest in energy.[33]

Table 1: Relative energies of the two lowest-lying Cobolt Phthalocyanine states calculated at CASSCF/cc-pVDZ and MRMP2/cc-pVDZ in different active spaces. Energies in eV. States are characterized by D_{4h} symmetry labels and their singly occupied metal MOs. The unoccupied $d_{x^2-y^2}$ orbital was omitted from reported calculations. Calculations using three *3d* metal orbitals rather than four omitted the fully occupied d_{xy} orbital.

Active Space		Calculations					
#occ	Metal	3	4	3	3	3	4
	Ligand	1	2	1	6	1	6
#virt	Ligand	2	3	4	2	2	4
	2nd set of metal	0	0	0	0	3	3
States and singly occupied metal MOs							
$^2A_{1g}$	d_{z^2}	0.0	0.0	0.0	0.0	0.0	0.0
^2E_g	d_{xz} / d_{yz}						
	ΔE (CASSCF)	0.60	0.60	0.60	0.60	0.46	0.57*
	ΔE (MRMP2)	0.36	0.35	0.36	0.33	0.28	0.32*

* This calculation was done at the CISDTQ level rather than CASSCF. i.e. orbitals were frozen and a maximum of 4 excitations was imposed, in order to reduce the computational memory requirements of the algorithm for the large active space. The occupied d_{xy} orbital was included in the active space, but its virtual counterpart was not.

The calculation in the fifth column included not only the highest energy occupied orbitals with predominantly *3d* metal character (see Figure 2) but also included

a second set of visually similar *3d* orbitals of predominantly d_{xz}, d_{yz}, and d_{z^2} character from the virtual orbitals. This lowered the difference in relative energies by 0.14 eV to 0.46 eV at the CASSCF level and reduced the MRMP2 energy difference by 0.08 eV to 0.28 eV.

The effect of including additional ligand orbitals in the active space can be seen by comparing columns 1 and 4, which led to a relative energy difference of 0.00 eV at CASSCF and 0.03 eV at MRMP2.

Ground state of FePc

Table 2 shows the results of different CASSCF and MRMP2 calculations on the lowest energy states of FePc using different active spaces. These calculations disagree as to whether the ground state is a $^3A_{2g}$ state or the ^3E_g degenerate pair. In the largest active space, the ^3E_g state is lowest energy by 0.13 eV after the perturbation theory correction is applied. However, both MRMP2 calculations on smaller active spaces, and the CASSCF calculations, suggest the $^3A_{2g}$ state is the lowest energy state, by around 0.02-0.05 eV. Figure 3 depicts the dominant electronic configuration of both these states. The next excited state of $^3B_{2g}$ symmetry can be ruled out as being the ground state as it is consistently found to be higher in energy by around 0.7 to 0.9 eV.

Table 2: Relative energies of the lowest-lying Iron Phthalocyanine states calculated at CASSCF/cc-pVDZ and MRMP2/cc-pVDZ in different active spaces. States are characterized by their D_{4h} symmetry label and by their singly occupied molecular orbitals (SOMOs). The unoccupied $d_{x^2-y^2}$ orbital was omitted from reported calculations. The calculation using three *3d* metal orbitals rather than four omitted the fully occupied d_{xy} orbital.

Active Space		Calculations					
#occ	Metal	4	4	3	4	4	4
	Ligand	1	1	1	2	1	2
#virt	Ligand	2	2	2	3	2	3
	2nd set of metal	0	0	3	0	4	4
#States weighted:		3	4	3	3	4	4
State and singly occupied metal MOs							
$^3A_{2g}$	d_{xz} and d_{yz}	0.0	0.0	0.0	0.0	0.0	0.0
^3E_g	d_{z^2} and one of d_{xz} / d_{yz}						
	ΔE (CASSCF)	0.023	-0.012	0.35	0.016	0.044	0.037
	ΔE (MRMP2)	0.055	0.018	0.38	0.055	-0.16	-0.13
$^3B_{2g}$	d_{xy} and d_{z^2}						
	ΔE (CASSCF)	0.97	0.88	n/a*	0.95	0.87	0.85
	ΔE (MRMP2)	0.95	0.85	n/a*	0.94	0.88	0.70

* this calculation's active space does not include the d_{xy} orbital required to describe this state.

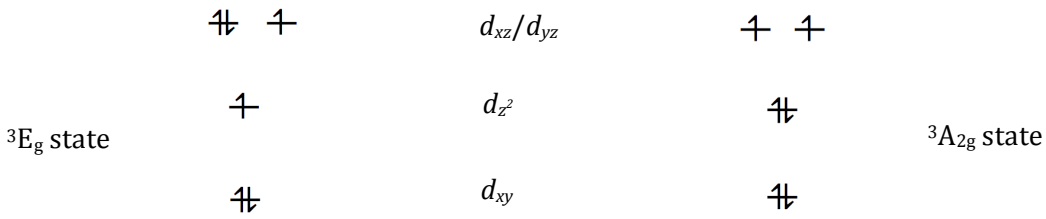


Figure 3: Two possibilities for the ground state configuration of FePc.

Both the $^3A_{2g}$ and $^3B_{2g}$ states are largely single reference, with a leading coefficient of 0.9 in the CASSCF wavefunction in the largest active space. Whereas the ^3E_g states have a leading coefficient of 0.84, meaning the dominant electronic configuration accounts for only around 70% of the total wave function. The ^3E_g states have a secondary contributing electronic configuration which accounts for a further 10% of the wave-function, in which the d_{xy} orbital is singly occupied as is one of the d_{xz}/d_{yz} orbitals.

Experimental studies that rely on a canonical square planar energy ordering for the 3d orbitals[27-31] uniformly predict a low-lying $^3B_{2g}$ state with orbital

occupancy $(d_{z^2})^1(d_{xz},d_{yz})^{2,2}(d_{xy})^1$ that is near-degenerate with a ^{3}Eg state $(d_{z^2})^2(d_{xz}, d_{yz})^{2,1}(d_{xy})^1$. Our results are in strong disagreement with those findings, and indicate the d_{xy} orbital is the lowest in energy and always doubly occupied. Our findings of a near-degeneracy between a $^3A_{2g}$ and fully-degenerate $^3E_{2g}$ states are consistent with other computational results[32,33,57] and experimental results including magnetic circular dichroism,[58] Mössbauer,[59] X-ray absorption and X-ray magnetic circular dichroism[54,60] spectra. For example, the Mössbauer spectroscopy[59] and magnetic susceptibility[61] measurements on crystalline α-FePc established an ^3E_g ground state, with occupied orbitals $(d_{xy})^2(d_{xz},d_{yz})^3(d_{z^2})^1$. In that instance the environmental effects in the condensed phased system may have broken the near-degeneracy between the $^3E_{2g}$ and $^3A_{2g}$ states.

Ground state of MnPc

Moving from FePc to MnPc, Figure 1 makes it clear that the d_{xy} orbital moves higher in energy to near the other *d*-orbitals. There are five electrons that need to be placed in these four *d*-orbitals of the Mn atom. After putting one electron into each of the four orbitals, a second electron must be paired in one of them. Table 3 shows the results from different CASSCF and MRMP2 calculations on MnPc.

Table 3: Possible MnPc ground states calculated at CASSCF/cc-pVDZ and MRMP2/cc-pVDZ. States are characterized by their D_{4h} symmetry labels and by which metal orbital is doubly occupied. The unoccupied $d_{x^2-y^2}$ orbital was omitted from reported calculations.

Active Space		Calculations					
#occ	Metal	4	4	4	4	4	4
	Ligand	1	2	2	2	1	2
#virt	Ligand	2	2	2	2	2	3
	2nd set of metal	0	0	0	0	4	4
#States weighted:		4	4	15	6	4	4
State and doubly-occupied metal MO							
$^4A_{2g}$	d_{xy}	0.0	0.0	0.0	0.0	0.0	0.0
^4E_g	d_{xz} / d_{yz}						
	ΔE (CASSCF)	0.51	0.50	0.56	0.61	0.45	0.44
	ΔE (MRMP2)	0.35	0.32	0.27	0.27	0.23	0.24
$^4B_{1g}$	d_z^2						
	ΔE (CASSCF)	0.55	0.53	0.59	0.65	0.40	0.40
	ΔE (MRMP2)	0.34	0.31	0.30	0.29	0.37	0.35

As Table 3 shows, all calculations suggest that the lowest energy state in the gas phase is the $^4A_{2g}$ state in which the d_{xy} orbital is doubly occupied and all other medal *d*-orbitals except $d_{x^2-y^2}$ are singly occupied. The degenerate ^4E_g states and the $^4B_{1g}$ state are all about 0.25 to 0.35 eV higher in energy according to the MRMP2 calculations. In all states the dominant electronic configuration in the largest CASSCF calculation has a coefficient of 0.91 meaning it accounts for about 83% of the total wave function in all states.

Most experimental evidence points to a $^4A_{2g}$ ground state for β-phase crystalline MnPc,[62-68] in agreement with our calculated results. In this phase, the crystal packing leads to axial interactions between manganese centers and the azamethine nitrogen atoms of adjacent molecules, which presumably destabilizes the out-of-plane d_{xy}/d_{yz} and d_{z^2} orbitals, and further favors the $^4A_{2g}$ state. However, experimental evidence regarding condensed phase systems with a significant dielectric response but without direct axial interactions (argon matrix,[10] thin films on gold substrate[3,4,54,69-71]) suggests aground state of ^4E_g for these systems rather than a $^4A_{2g}$ state. The environment of these systems may be substantially favoring the out-of-plane d_{xy}/d_{yz} orbitals, swapping the energy ordering of the ^4E_g and $^4A_{2g}$ states. Alternatively it is possible that the limited size

of the active spaces in our calculations has unduly prejudiced the $^4A_{2g}$ state over the ^4E_g states, and with a larger active space we might see these reordered.

Conclusions

For CoPc all our calculations predicted an $^2A_{1g}$ ground state, with occupation of $(d_{xy})^2(d_{xz},d_{yz})^4(d_{z^2})^1(d_{x^2-y^2})^0$ for the metal *3d*-orbitals, i.e. the d_{z^2} orbital was singly occupied (the $d_{x^2-y^2}$ orbital is not occupied at all for any of CoPc, FePc, or MnPc). The ^2E_g state was about 0.3 eV higher in energy at the MRMP2 level with an occupation of $(d_{xy})^2(d_{xz},d_{yz})^3(d_{z^2})^2(d_{x^2-y^2})^0$ for the metal *3d*-orbitals.

For FePc our MRMP2 calculations in the largest active space gave a degenerate ground state pair of symmetry ^3E_g, giving a *3d* metal orbital occupation of $(d_{xy})^2(d_{xz},d_{yz})^3(d_{z^2})^1(d_{x^2-y^2})^0$. However a $^3A_{2g}$ state with occupation $(d_{xy})^2(d_{xz},d_{yz})^2(d_{z^2})^2(d_{x^2-y^2})^0$ was found to be very similar in energy, with most calculations suggesting it was within about 0.1 eV of the degenerate pair. The $^3B_{2g}$ state with occupation $(d_{xy})^1(d_{xz},d_{yz})^4(d_{z^2})^1(d_{x^2-y^2})^0$ was found to be 0.7 eV higher in energy in the largest active space.

For MnPc, the $^4A_{2g}$ state with occupation $(d_{xy})^2(d_{xz},d_{yz})^2(d_{z^2})^1(d_{x^2-y^2})^0$ was consistently the lowest in energy in all calculations. The ^4E_g and $^4B_{1g}$ states were found to be about 0.25 to 0.35 eV higher in energy, with occupations of $(d_{xy})^1(d_{xz},d_{yz})^3(d_{z^2})^1(d_{x^2-y^2})^0$ and $(d_{xy})^1(d_{xz},d_{yz})^2(d_{z^2})^2(d_{x^2-y^2})^0$ respectively. Calculations varied on the ordering of these higher energy states, but the calculation in the largest active space found that the ^4E_g pair of states was lower than the $^4B_{1g}$ state by about 0.1 eV.

In the calculations in this chapter, the MRMP2 method yielded more consistent results than in the calculations on the ZnPc spectrum in the previous chapter. The perturbation theory correction rarely altered the relative energies of the CASSCF states by more than 0.2 eV. This can be compared to the previous chapter where relative energy corrections were in the range of 1 to 3 eV. This

can be attributed to the inclusion of sufficient orbitals in the CASSCF active space. These calculations, focused on the occupation of the *3d*-metal orbitals, were able to be performed with a complete set of the relevant *3d* orbitals included, as well as additional virtual copies of them, whereas in the previous chapter it was only possible to include a small subset of the ligand π-orbitals in the active space.

References

(1) Ahlund, J.; Nilson, K.; Schiessling, J.; Kjeldgaard, L.; Berner, S.; Martensson, N.; Puglia, C.; Brena, B.; Nyberg, M.; Luo, Y. *J. Chem. Phys.* **2006**, *125*.

(2) Bialek, B.; Kim, I. G.; Lee, J. I. *Surf Sci* **2003**, *526*, 367.

(3) Brumboiu, I. E.; Totani, R.; de Simone, M.; Coreno, M.; Grazioli, C.; Lozzi, L.; Herper, H. C.; Sanyal, B.; Erikssont, O.; Puglia, C.; Brena, B. *J. Phys. Chem. A* **2014**, *118*, 927.

(4) Grobosch, M.; Mahns, B.; Loose, C.; Friedrich, R.; Schmidt, C.; Kortus, J.; Knupfer, M. *Chem. Phys. Lett.* **2011**, *505*, 122.

(5) Kataoka, T.; Sakamoto, Y.; Yamazaki, Y.; Singh, V. R.; Fujimori, A.; Takeda, Y.; Ohkochi, T.; Fujimori, S. I.; Okane, T.; Saitoh, Y.; Yamagami, H.; Tanaka, A. *Solid State Commun* **2012**, *152*, 806.

(6) Kroll, T.; Kraus, R.; Schonfelder, R.; Aristov, V. Y.; Molodtsova, O. V.; Hoffmann, P.; Knupfer, M. *J. Chem. Phys.* **2012**, *137*, 054306.

(7) Liao, M. S.; Watts, J. D.; Huang, M. J. *Inorg Chem* **2005**, *44*, 1941.

(8) Marom, N.; Kronik, L. *Applied Physics A - Materials Science & Processing* **2009**, *95*, 165.

(9) Wang, J. H.; Shi, Y. S.; Cao, J. X.; Wu, R. Q. *App. Phys. Lett.* **2009**, *94*.

(10) Williamson, B. E.; Vancott, T. C.; Boyle, M. E.; Misener, G. C.; Stillman, M. J.; Schatz, P. N. *J Am Chem Soc* **1992**, *114*, 2412.

(11) Ueno, L. T.; Jayme, C. C.; Silva, L. R.; Pereira, E. B.; de Oliveira, S. M.; Machado, A. E. H. *J Brazil Chem Soc* **2012**, *23*, 2237.

(12) Vancott, T. C.; Rose, J. L.; Misener, G. C.; Williamson, B. E.; Schrimpf, A. E.; Boyle, M. E.; Schatz, P. N. *J. Phys. Chem.* **1989**, *93*, 2999.

(13) Schmidt, M. W.; Gordon, M. S. *Annu Rev Phys Chem* **1998**, *49*, 233.

(14) Szalay, P. G.; Muller, T.; Gidofalvi, G.; Lischka, H.; Shepard, R. *Chem. Rev.* **2012**, *112*, 108.

(15) Hirao, K. *Chem. Phys. Lett.* **1992**, *196*, 397.

(16) Hirao, K. *Chem. Phys. Lett.* **1992**, *190*, 374.

(17) Hirao, K. *Int J Quantum Chem* **1992**, 517.

(18) Nakano, H. *J. Chem. Phys.* **1993**, *99*, 7983.

(19) Nakano, H. *Chem. Phys. Lett.* **1993**, *207*, 372.

(20) Balabanov, N. B.; Peterson, K. A. *J. Chem. Phys.* **2005**, *123*.

(21) Balabanov, N. B.; Peterson, K. A. *J. Chem. Phys.* **2006**, *125*.

(22) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(23) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(24) Gordon, M. S.; Schmidt, M.W. In *Theory and Applications of Computational Chemistry*; Dykstra, C.E.; Frenking, G.; Kim, K.S.; Scuseria, G.E., Ed.; Elsevier: Amsterdam, 2005.

(25) Bode, B. M.; Gordon, M. S. *J Mol Graph Model* **1998**, *16*, 133.

(26) Pulay, P. *Int J Quantum Chem* **2011**, *111*, 3273.

(27) Barraclough, C.G.; Martin, R. L.; Mitra, S.; Sherwood, R. C. *J. Chem. Phys.* **1970**, *53*, 1643.

(28) Miedema, P. S.; Stepanow, S.; Gambardella, P.; de Groot, F. M. F. *J Phys Conf Ser* **2009**, *190*.

(29) Stepanow, S.; Miedema, P. S.; Mugarza, A.; Ceballos, G.; Moras, P.; Cezar, J. C.; Carbone, C.; de Groot, F. M. F.; Gambardella, P. *Phys Rev B* **2011**, *83*.

(30) Kuz'min, M. D.; Savoyant, A.; Hayn, R. *J. Chem. Phys.* **2013**, *138*.

(31) Fernandez-Rodriguez, J.; Toby, B.; Van Veenendaal, M. *Phys Rev B* **2015**, *91*.

(32) Bruder, I.; Schoneboom, J.; Dinnebier, R.; Ojala, A.; Schafer, S.; Sens, R.; Erk, P.; Weis, J. *Org Electron* **2010**, *11*, 377.

(33) Liao, M. S.; Scheiner, S. *J. Chem. Phys.* **2001**, *114*, 9780.

(34) Garcia-Lastra, J. M.; Cook, P. L.; Himpsel, F. J.; Rubio, A. *J. Chem. Phys.* **2010**, *133*.

(35) Theisen, R. F.; Huang, L.; Fleetham, T.; Adams, J. B.; Li, J. *J. Chem. Phys.* **2015**, *142*.

(36) Ricciardi, G.; Rosa, A.; Baerends, E. J. *J. Phys. Chem. A* **2001**, *105*, 5242.

(37) Plows, F. L.; Jones, A. C. *J Mol Spectrosc* **1999**, *194*, 163.

(38) Peralta, G. A.; Seth, M.; Zhekova, H.; Ziegler, T. *Inorg Chem* **2008**, *47*, 4185.

(39) Peralta, G. A.; Seth, M.; Ziegler, T. *Inorg Chem* **2007**, *46*, 9111.

(40) Nguyen, K. A.; Pachter, R. *J. Chem. Phys.* **2001**, *114*, 10757.

(41) Nemykin, V. N.; Hadt, R. G.; Belosludov, R. V.; Mizuseki, H.; Kawazoe, Y. *J. Phys. Chem. A* **2007**, *111*, 12901.

(42) Mack, J.; Stillman, M. J. *J. Phys. Chem.* **1995**, *99*, 7935.

(43) Krasnikov, S. A.; Preobrajenski, A. B.; Sergeeva, N. N.; Brzhezinskaya, M. M.; Nesterov, M. A.; Cafolla, A. A.; Senge, M. O.; Vinogradov, A. S. *Chem. Phys.* **2007**, *332*, 318.

(44) Guo, M. Y.; He, R. X.; Dai, Y. L.; Shen, W.; Li, M.; Zhu, C. Y.; Lin, S. H. *J. Chem. Phys.* **2012**, *136*.

(45) Gantchev, T. G.; van Lier, J. E.; Hunting, D. J. *Radiat Phys Chem* **2005**, *72*, 367.

(46) Evangelista, F.; Carravetta, V.; Stefani, G.; Jansik, B.; Alagia, M.; Stranges, S.; Ruocco, A. *J. Chem. Phys.* **2007**, *126*.

(47) Cook, P. L.; Yang, W. L.; Liu, X. S.; Garcia-Lastra, J. M.; Rubio, A.; Himpsel, F. J. *J. Chem. Phys.* **2011**, *134*.

(48) Schwieger, T.; Peisert, H.; Golden, M. S.; Knupfer, M.; Fink, J. *Phys Rev B* **2002**, *66*.

(49) Dunford, C. L.; Williamson, B. E. *J. Phys. Chem. A* **1997**, *101*, 2050.

(50) Henriksson, A.; Sundbom, M.; Roos, B. *Theor Chim Acta* **1972**, *27*, 303.

(51) Marom, N.; Hod, O.; Scuseria, G. E.; Kronik, L. *J. Chem. Phys.* **2008**, *128*.

(52) Marom, N.; Ren, X. G.; Moussa, J. E.; Chelikowsky, J. R.; Kronik, L. *Phys Rev B* **2011**, *84*.

(53) Downes, J. E.; McGuinness, C.; Glans, P. A.; Learmonth, T.; Fu, D. F.; Sheridan, P.; Smith, K. E. *Chem. Phys. Lett.* **2004**, *390*, 203.

(54) Kroll, T.; Kraus, R.; Schonfelder, R.; Aristov, V. Y.; Molodtsova, O. V.; Hoffmann, P.; Knupfer, M. *J. Chem. Phys.* **2012**, *137*.

(55) Janczak, J.; Kubiak, R. *Inorg Chim Acta* **2003**, *342*, 64.

(56) Reynolds, P. A.; Figgis, B. N. *Inorg Chem* **1991**, *30*, 2294.

(57) Liao, M. S.; Scheiner, S. *J. Chem. Phys.* **2002**, *116*, 3635.

(58) Stillman, M. J.; Thomson, A. J. *J Chem Soc Farad T 2* **1974**, *70*, 790.

(59) Filoti, G.; Kuz'min, M. D.; Bartolome, J. *Phys Rev B* **2006**, *74*.

(60) Kuz'min, M. D.; Hayn, R.; Oison, V. *Phys Rev B* **2009**, *79*.

(61) Dale, B. W.; Williams, R. J. P.; Edwards, P. R.; Johnson, C. E. *T Faraday Soc* **1968**, *64*, 620.

(62) Barraclough, C. G.; Gregson, A. K.; Mitra, S. *The Journal of Chemical Physics* **1974**, *60*, 962.

(63) Mitra, S.; Gregson, A. K.; Hatfield, W. E.; Weller, R. R. *Inorg Chem* **1983**, *22*, 1729.

(64) Miyoshi, H. *B Chem Soc Jpn* **1974**, *47*, 561.

(65) Turner, P.; Gunter, M. J. *Inorg Chem* **1994**, *33*, 1406.

(66) Barraclough, C.G.; Martin, R. L.; Mitra, S.; Sherwood, R. C. *J. Chem. Phys.* **1970**, *53*, 1638.

(67) Labarta, A.; Molins, E.; Tejada, J. *Z Phys B Con Mat* **1985**, *58*, 299.

(68) Lever, A. B. P. *J Chem Soc* **1965**, 1821.

(69) Grobosch, M.; Schmidt, C.; Kraus, R.; Knupfer, M. *Org Electron* **2010**, *11*, 1483.

(70) Kraus, R.; Grobosch, M.; Knupfer, M. *Chem. Phys. Lett.* **2009**, *469*, 121.

(71) Petraki, F.; Peisert, H.; Hoffmann, P.; Uihlein, J.; Knupfer, M.; Chasse, T. *J. Phys. Chem. C* **2012**, *116*, 5121.

Conclusions

Accounting for electron correlation remains the largest outstanding problem in computational quantum chemistry. Chapter 2 demonstrated that some progress can be made through treating static and dynamic correlation separately, as they exhibit different convergence behavior towards the one-electron complete basis set limit. It was observed empirically that, like HF energies, static correlation energies converge towards the CBS limit at a root-exponential rate as a function of basis set cardinality for correlation-consistent and polarization-consistent basis sets. As a result of this rapid convergence, static correlation energies were within chemical accuracy for all test molecules in our dataset for the cc-pVTZ basis and above. It was also observed that the 6-31G(*2p*,*2d*) basis gave good results for its small size, averaging results within chemical accuracy across the dataset.

The study of the dicationic dimer cyclopropenium system in chapter 3 showed that dynamic correlation / dispersion can be sufficiently strong to hold together what would otherwise be an electrostatically repulsive system. Long range interactions with the counterions provided necessary environmental charge balance to stabilize the overall system and hold it together electrostatically, as the system is only metastable in the gas phase. However, it is remarkable that within the charge-balanced ionic-liquid environment the cations should seek each other out and form dications rather than spreading apart from each other with an anion between each cation. This demonstrates the strength and importance of dispersion in intermolecular interactions, contrary to conventional textbook wisdom.

In chapter 4 it was demonstrated that the zinc phthalocyanine spectrum can be reproduced using EOM-CCSD and EOM-CCSD(T) methods. Due to its extensive macrocycle system, this molecule has a number of energetically-close occupied π-orbitals, and the excited states involving transitions from these orbitals are strongly multireference, and thus static correlation plays a major role in the transition energies.

In chapter 5 it was observed that CASSCF-based methods could not be used to predict electronic excitation energies due to practical limitations on the size of the active space. The number of π and π* ligand orbitals in the molecule proved too large for these method to reliably capture the static correlation energy. The perturbation theory energy corrections showed numerical instabilities between different calculations when applied to more than one excited state.

Chapter 6 explored the ground states of Mn, Fe and Co phthalocyanine, which are open-shell systems whose ground states are not unanimously agreed upon in the literature. Using CASSCF and MRMP2 methods the ground states of the Co, Fe, and Mn phthalocyanine systems were assigned as $^2A_{1g}$, ^3E_g, and $^4A_{2g}$ respectively, with *3d*-metal orbital occupancies of $(d_{xy})^2(d_{xz},d_{yz})^4(d_{z^2})^1(d_{x^2-y^2})^0$, $(d_{xy})^2(d_{xz},d_{yz})^3(d_{z^2})^1(d_{x^2-y^2})^0$, and $(d_{xy})^2(d_{xz},d_{yz})^2(d_{z^2})^1(d_{x^2-y^2})^0$ respectively.

Future work

The main unresolved issue arising from this work concerns the assignment of the absorption spectra of the other metallophthalocyanines, particularly FePc and MnPc in which the *3d*-orbitals become near-degenerate with each other and the ligand HOMO and LUMO. These systems have quite different looking electronic absorption spectra to ZnPc, and have not been assigned by any previous computational work.

The work in chapter 4 on the electronic excitation spectrum of ZnPc suggests it should be possible to use EOM-CCSD based methods to predict the spectra of the open-shell phthalocyanines. For some of the open-shell phthalocyanines, it is non-trivial to obtain MOs to use in the EOM-CCSD calculation. HF calculations for the open-shell metallophthalocyanines often do not converge unless the initial guess is well chosen. CASSCF calculations on these systems converge reliably, however, and provide the easiest way of obtaining molecular orbitals that can be used in HF or EOM-CCSD calculations.

Exploratory CASSCF calculations on MnPc suggested that many of the spectral transitions might be double-excitations and very multireference in nature. Any program that utilizes a CIS initial guess for finding excited states will likely not find these roots. Furthermore, the EOM-CCSD method is known to yield energies inaccurate by up to 2 eV for double-excitations, and thus the use of the EOM-CCSD(T) method would likely be required to enable confident assignment of the experimental spectrum.

Appendix 1: Supporting Information for Chapter 2

Optimal Composition of Atomic Orbital Basis Sets for Recovering Static Correlation Energy

Table 1: Geometries of the molecules used

BeH			
BE	0.00000	0.00000	0.00000
H	0.00000	0.00000	1.34260
CH			
C	0.00000	0.00000	0.00000
H	0.00000	0.00000	1.11990
CH2 (1A1)			
C	0.00000	0.00000	0.17349
H	0.86312	0.00000	-0.52047
H	-0.86312	0.00000	-0.52047
CH2 (3B1)			
C	0.00000	0.00000	0.10534
H	0.98877	0.00000	-0.31600
H	-0.98877	0.00000	-0.31600
CH3			
C	0.00000	0.00000	0.00000
H	1.07610	0.00000	0.00000
H	-1.46998	0.00000	
H	-0.53805	0.93193	0.00000
CH3CH2 - transition state			
C	-0.25872	-0.81683	0.00000
C	-0.25099	0.67419	0.00000
H	0.75883	-1.22594	0.00000
H	-0.75883	-1.21387	0.88342
H	-0.75883	-1.21387	-0.88342
H	-0.17002	1.22594	-0.92432
H	-0.17002	1.22594	0.92432
CH3Cl			
C	0.00000	0.00000	1.12100
CL	0.00000	0.00000	-0.65460
H	1.03055	0.00000	1.46737
H	-0.51528	0.89248	1.46737
H	-0.51528	-0.89248	1.46737
CH3F			
C	-0.63207	0.00000	0.00000
F	0.74912	0.00000	0.00000
H	-0.98318	-0.33849	0.97263

H	-0.98322	1.01155	-0.19317
H	-0.98320	-0.67308	-0.77944
CH3FCl -- transition state			
CL	1.45475	-0.00124	-0.00004
F	-0.32359	0.00463	0.00012
C	-2.38742	-0.00215	-0.00007
H	-2.49509	-0.85536	-0.64940
H	-2.49731	-0.13867	1.06314
H	-2.50154	0.98627	-0.41373
CH3SH			
C	-0.04775	1.15011	0.00000
S	-0.04775	-0.66364	0.00000
H	1.27806	-0.82966	0.00000
H	-1.09120	1.45643	0.00000
H	0.43178	1.54538	0.89100
H	0.43178	1.54538	-0.89100
CH4			
C	0.00000	0.00000	0.00000
H	0.62760	0.62760	0.62760
H	0.62760	-0.62760	-0.62760
H	-0.62760	0.62760	-0.62760
H	-0.62760	-0.62760	0.62760
Cl2			
CL	0.00000	0.00000	0.00000
CL	0.00000	0.00000	1.98790
ClCH3Cl- -- transition state			
CL	0.00003	0.01953	2.32250
C	0.00051	0.00049	-0.00009
H	0.76128	-0.75073	0.00638
H	-1.03045	-0.28272	0.00215
H	0.27073	1.03493	-0.00870
CL	-0.00030	-0.01978	-2.32246
ClF			
F	0.00000	0.00000	0.00000
CL	0.00000	0.00000	1.62830
ClO			
O	0.00000	0.00000	0.00000
CL	0.00000	0.00000	1.56960
CN			
C	0.00000	0.00000	0.00000
N	0.00000	0.00000	1.17180
CO			
C	0.00000	0.00000	0.00000
O	0.00000	0.00000	1.12830
CO2			
C	0.00000	0.00000	0.00000

O	0.00000	0.00000	1.15996
O	0.00000	0.00000	-1.15996
CS			
C	0.00000	0.00000	0.00000
S	0.00000	0.00000	1.53490
F2			
F	0.00000	0.00000	0.00000
F	0.00000	0.00000	1.41190
FCH3Cl- -- transition state			
F	0.00000	0.00000	-2.53793
C	0.00000	0.00000	-0.48837
H	1.06209	0.00000	-0.61497
H	-0.53104	0.91979	-0.61497
H	-0.53104	-0.91979	-0.61497
CL	0.00000	0.00000	1.62450
H2			
H	0.00000	0.00000	0.00000
H	0.00000	0.00000	0.74140
H2CCH2			
C	0.66559	0.00000	0.00000
C	-0.66559	0.00000	0.00000
H	1.23167	0.92150	0.00000
H	1.23167	-0.92150	0.00000
H	-1.23167	0.92150	0.00000
H	-1.23167	-0.92150	0.00000
H2CO			
O	0.00000	0.00000	0.67462
C	0.00000	0.00000	-0.52971
H	0.00000	0.93549	-1.10937
H	0.00000	-0.93549	-1.10937
H2NNH2			
N	0.70986	-0.10034	-0.07513
N	-0.70986	0.10034	-0.07513
H	1.04953	-0.37288	0.83834
H	1.13826	0.78185	-0.31241
H	-1.04953	0.37288	0.83834
H	-1.13826	-0.78185	-0.31241
H2O			
O	0.00000	0.00000	0.00000
H	0.00000	0.75751	0.58717
H	0.00000	-0.75751	0.58717
H2S			
S	0.00000	0.00000	0.00000
H	0.00000	0.96160	0.92690
H	0.00000	-0.96160	0.92690
H3CCH3			

C	-0.76309	0.00000	0.00000
C	0.76309	0.00000	0.00000
H	1.15831	-0.44364	-0.91410
H	-1.15831	-1.01333	-0.07464
H	1.15831	-0.56982	0.84125
H	1.15831	1.01346	0.07285
H	-1.15831	0.44202	0.91488
H	-1.15831	0.57130	-0.84024
H3COH			
C	-0.04642	0.66307	0.00000
O	-0.04642	-0.75506	0.00000
H	-1.08696	0.97594	0.00000
H	0.86059	-1.05704	0.00000
H	0.43815	1.07159	0.88954
H	0.43815	1.07159	-0.88954
HC2H4 - transition state			
C	-0.56788	0.00005	-0.21896
C	0.75114	-0.00004	0.04193
H	-1.49388	-0.00049	1.53177
H	-1.10169	0.92065	-0.40863
H	-1.10202	-0.92023	-0.40911
H	1.29913	-0.92234	0.17376
H	1.29890	0.92233	0.17436
HCCH			
C	0.00000	0.00000	-0.60157
C	0.00000	0.00000	0.60157
H	0.00000	0.00000	1.66418
H	0.00000	0.00000	-1.66418
HCl			
CL	0.00000	0.00000	0.00000
H	0.00000	0.00000	1.27460
HClH -- transition state			
H	0.00000	0.00000	1.48580
CL	0.00000	0.00000	0.00000
H	0.00000	0.00000	-1.48580
HCN			
C	0.00000	0.00000	0.00000
H	0.00000	0.00000	1.06501
N	0.00000	0.00000	-1.15324
HCN -- transition state			
C	0.08032	0.62026	0.00000
N	0.08032	-0.56810	0.00000
H	-1.04415	0.25512	0.00000
HCO			
C	0.00000	0.00000	0.00000
H	1.11910	0.00000	0.00000

O	-0.66460	0.96950	0.00000
HF			
F	0.00000	0.00000	0.00000
H	0.00000	0.00000	0.91680
HN2			
N	-0.06244	0.65949	0.00000
N	-0.06244	-0.51871	0.00000
H	0.87419	-0.98548	0.00000
HN2 -- transition state			
N	0.08456	-0.64293	0.00000
N	0.08456	0.47988	0.00000
H	-1.18388	1.14140	0.00000
** HN2O -- transition state			
H	-0.30329	-1.93071	0.00000
O	-0.86101	-0.62153	0.00000
N	0.00000	0.25703	0.00000
N	1.02733	0.72910	0.00000
HNC			
C	0.00000	0.00000	-0.73725
N	0.00000	0.00000	0.43209
H	0.00000	0.00000	1.42696
HOCH3F- -- transition state			
F	1.85061	-0.01318	-0.00013
C	0.09086	0.01059	0.00027
H	0.04091	1.07955	-0.01175
H	0.03716	-0.52801	-0.92294
H	0.03749	-0.50746	0.93513
O	-1.89280	0.10327	-0.00012
H	-2.17382	-0.81511	0.00004
HOCH4 -- transition state			
C	-1.21149	0.00797	0.00041
O	1.29397	-0.10869	0.00013
H	0.00948	-0.11802	0.00280
H	-1.52553	-0.23325	1.01007
H	-1.43067	1.03323	-0.27808
H	-1.55271	-0.71011	-0.73770
H	1.41664	0.84989	-0.00059
HOCl			
O	1.09185	-0.11725	0.00000
H	1.36860	0.80575	0.00000
CL	-0.59432	0.00778	0.00000
HOH -- transition state			
H	0.00000	0.00000	-0.86029
O	0.00000	0.00000	0.32902
H	0.00000	0.00000	-1.77191
Li2			

LI	0.00000	0.00000	0.00000
LI	0.00000	0.00000	2.67300
LiF			
F	0.00000	0.00000	0.00000
LI	0.00000	0.00000	1.56390
LiH			
LI	0.00000	0.00000	0.00000
H	0.00000	0.00000	1.59570
N2			
N	0.00000	0.00000	0.54880
N	0.00000	0.00000	-0.54880
N2O			
N	0.00000	0.00000	-1.19567
N	0.00000	0.00000	-0.07511
O	0.00000	0.00000	1.11194
Na2			
NA	0.00000	0.00000	0.00000
NA	0.00000	0.00000	3.07890
NaCl			
NA	0.00000	0.00000	0.00000
CL	0.00000	0.00000	2.36080
NH			
N	0.00000	0.00000	0.00000
H	0.00000	0.00000	1.03620
NH2			
N	0.00000	0.00000	0.14190
H	-0.80178	0.00000	-0.49666
H	0.80178	0.00000	-0.49666
NH3			
N	0.00000	0.00000	0.11289
H	0.00000	0.93802	-0.26341
H	0.81235	-0.46901	-0.26341
H	-0.81235	-0.46901	-0.26341
O2			
O	0.00000	0.00000	0.00000
O	0.00000	0.00000	1.20750
O3			
O	0.00000	0.00000	0.44434
O	1.08340	0.00000	-0.22217
O	-1.08340	0.00000	-0.22217
OH-			
O	0.00000	0.00000	0.10689
H	0.00000	0.00000	-0.85515
OH			
O	0.00000	0.00000	0.00000
H	0.00000	0.00000	0.96970

P2

P	0.00000	0.00000	0.00000
P	0.00000	0.00000	1.89340

PH2

P	0.00000	0.00000	-0.11566
H	1.02013	0.00000	0.86743
H	-1.02013	0.00000	0.86743

PH3

P	0.00000	0.00000	0.12752
H	1.18792	0.00000	-0.63760
H	-0.59396	-1.02877	-0.63760
H	-0.59396	1.02877	-0.63760

S2

S	0.00000	0.00000	0.00000
S	0.00000	0.00000	1.88920

SH

S	0.00000	0.00000	0.07884
H	0.00000	0.00000	-1.26137

SH3 -- transition state

H	1.26210	-0.22010	0.00000
S	0.00000	0.22315	0.00000
H	-0.50058	-1.11545	0.00000
H	-0.76152	-2.23491	0.00000

Si2

SI	0.00000	0.00000	
SI	0.00000	2.24600	

Si2H6

SI	0.00000	0.00000	1.17031
SI	0.00000	0.00000	-1.17031
H	1.38811	0.00000	1.68549
H	-0.69406	1.20214	1.68549
H	-0.69406	-1.20214	1.68549
H	-1.38811	0.00000	-1.68549
H	0.69406	1.20214	-1.68549
H	0.69406	-1.20214	-1.68549

SiH2 (1A1)

SI	0.00000	0.00000	0.13097
H	0.00000	1.09475	-0.91676
H	0.00000	-1.09475	-0.91676

SiH2 (3B1)

SI	0.00000	0.00000	0.00000
H	0.00000	0.00000	-1.51400
H	0.00000	1.51300	0.05550

SiH3

SI	0.00000	0.00000	0.07918
H	0.00000	1.40798	-0.36949

H	1.21935	-0.70399	-0.36949
H	-1.21935	-0.70399	-0.36949
SiH4			
SI	0.00000	0.00000	0.00000
H	0.85258	0.85258	0.85258
H	-0.85258	-0.85258	0.85258
H	-0.85258	0.85258	-0.85258
H	0.85258	-0.85258	-0.85258
SiO			
SI	0.00000	0.00000	0.00000
O	0.00000	0.00000	1.50970
SO			
S	0.00000	0.00000	0.00000
O	0.00000	0.00000	1.48110
SO2			
S	0.00000	0.00000	0.00000
O	0.00000	1.23490	0.72264
O	0.00000	-1.23490	0.72264

Table 2: Linear Regression Results. ΔE_{stat} as a function of number of valence electrons

	Intercept	95%CI	Slope	95%CI	R^2
pc-0[sp]	0.002375149	0.00315964	0.000255171	0.00021434	0.06556194
pc-1[sp]	5.15361E-05	0.003023571	0.000499053	0.000205109	0.226644391
pc-2[sp]	-0.000268737	0.003306991	0.000475645	0.000224336	0.18203292
pc-3[sp]	-0.000330026	0.00314602	0.000483807	0.000213416	0.202812412
pc-4[sp]	-0.000244376	0.003114918	0.000469806	0.000211306	0.196601857
pc-0[spd_2]	4.51056E-05	0.001689958	0.000344321	0.000114641	0.308708766
pc-1[spd] = pc-1	0.001066711	0.000782219	1.04465E-05	5.30632E-05	0.001914985
pc-1[spd_2]	0.000687798	0.000454668	2.07984E-05	3.08432E-05	0.022014829
pc-1[spd_3]	0.000497169	0.000350664	2.71698E-05	2.37879E-05	0.060663017
pc-2[spd_1}]	0.0009076	0.000805725	2.66104E-05	5.46578E-05	0.011597856
pc-2[spd]	0.000413286	0.000387094	2.62439E-05	2.62592E-05	0.047116804
pc-2[spd_3]	0.00031546	0.000220165	1.71575E-05	1.49353E-05	0.06132539
pc-3[spd_2]	0.000314678	0.000362798	1.1102E-05	2.46111E-05	0.009973126
pc-3[spd]	2.40112E-05	0.000156988	1.73508E-05	1.06496E-05	0.116144462
pc-4[spd]	1.45633E-05	0.000142518	1.5913E-05	9.67E-06	0.118255757
pc-2[spdf] = pc-2	0.000527054	0.000307089	8.78E-06	2.08319E-05	0.008720192
pc-3[spdf]	4.58671E-05	3.15971E-05	1.02E-07	2.14E-06	0.000113076
pc-4[spdf]	3.27562E-05	2.7457E-05	5.09E-07	1.86E-06	0.003678496
pc-3[spdfg] = pc-3	3.91371E-05	3.00127E-05	8.47E-07	2.04E-06	0.00849605
pc-4[spdfg] = pc-4	2.27179E-05	2.85761E-05	1.28E-06	1.94E-06	0.021164084
STO-3G	-0.006031233	0.00655004	0.001296702	0.000444334	0.296569776
6-31G	-0.000403411	0.00325133	0.000482325	0.00022056	0.191422376
6-311G	-0.000551204	0.003219105	0.000525641	0.000218374	0.222895155
6-31G(d,p)	0.000498121	0.000831201	5.90603E-05	5.6386E-05	0.051513743
6-311G(d,p)	-8.79818E-05	0.000875104	9.7284E-05	5.93643E-05	0.117345675
6-31G(2d,2p)	0.000174452	0.000483239	5.8057E-05	3.27814E-05	0.134404482
6-311G(2d,2p)	-9.23154E-05	0.000343255	4.9891E-05	2.32853E-05	0.185177114
6-311G(2df,2pd)	-4.60946E-05	0.00022267	3.03696E-05	1.51052E-05	0.166742986
cc-pVDZ[sp]	-0.000255666	0.00311647	0.000491362	0.000211411	0.210994286
cc-pVTZ[sp]	-0.000295952	0.003256847	0.00048558	0.000220934	0.192984483
cc-pVQZ[sp]	-0.000342682	0.003177843	0.000499295	0.000215575	0.209836214
cc-pV5Z[sp]	-0.00035923	0.003152409	0.000485758	0.000213849	0.203458809
cc-pVDZ[spd] = cc-pVDZ	2.53693E-05	0.000618293	8.34841E-05	4.1943E-05	0.163966821
cc-pVTZ[spd]	6.95761E-05	0.000231388	2.86485E-05	1.56966E-05	0.141561416
cc-pVQZ[spd]	3.38E-06	0.000181795	2.07544E-05	1.23324E-05	0.122966956
cc-pV5Z[spd]	8.50E-06	0.000166965	1.87282E-05	1.13264E-05	0.119212374
cc-pVTZ[spdf] = cc-pVTZ	0.000152589	0.000137063	1.01367E-05	9.30E-06	0.055569412
cc-pVQZ[spdf]	-4.69E-06	2.85083E-05	4.08E-06	1.93E-06	0.180203573
cc-pV5Z[spdf]	2.98E-07	2.10E-06	7.55E-08	1.43E-07	0.013690629
cc-pVQZ[spdfg] = cc-pVQZ	-4.18E-06	2.60551E-05	3.37E-06	1.77E-06	0.152362207

Table 3: CBS extrapolation parameters for static correlation energies calculated using correlation consistent basis sets

	-Es_infty	A	B
al	0.024146184	-0.303128918	2.5
b	0.03473193	-0.591292197	2.5
be	0.043816089	-0.454175441	2.5
beh	0.027655192	-0.535296131	2.5
c	0.019291065	-0.278080082	2.5
ch	0.043082793	-0.462032818	2.5
ch2_sing	0.062072981	-0.790339225	2.5
ch2_trip	0.038756192	-0.402102929	2.5
ch3	0.059384849	-0.564609957	2.5
ch3ch2	0.128940863	-1.098825973	2.5
ch3cl	0.087229672	-0.836097963	2.5
ch3f	0.093812542	-0.846667564	2.5
ch3fcl	0.112293667	-0.096099706	2.5
ch3sh	0.104686059	-0.915313801	2.5
ch4	0.08317673	-0.943153396	2.5
cl2	0.023350766	0.133496467	2.5
clch3cl-	0.0802183	-0.267609748	2.5
clf	0.037727454	0.251198104	2.5
clo	0.042645629	-0.05635818	2.5
cn	0.151911312	-0.13157038	2.5
co	0.131771174	-0.718304389	2.5
co2	0.176354295	-0.149856993	2.5
cs	0.104576329	-0.33612142	2.5
f2	0.079128721	0.031292753	2.5
fch3cl-	0.077140773	-0.484824058	2.5
h2	0.018542888	-0.125433727	2.5
h2cch2	0.144175731	-0.708041831	2.5
h2co	0.147075835	-2.203248442	2.5
h2nnh2	0.13504683	-1.558269011	2.5
h2o	0.053744116	-0.673979522	2.5
h2s	0.034435726	-0.289487187	2.5
h3cch3	0.151160148	-1.431992985	2.5
h3coh	0.118783498	-1.238699337	2.5
hc2h4	0.148371699	-0.703140769	2.5
hcch	0.148040605	-0.515815756	2.5
hcl	0.017368724	-0.155226548	2.5
hclh	0.037438449	0.274732723	2.5
hcn	0.151429758	-1.054288795	2.5
hcn_ts	0.141869871	-0.461268125	2.5
hco	0.125828392	-0.477726244	2.5
hf	0.024456445	-0.226102036	2.5
hn2	0.129970892	-0.102576626	2.5
hn2_ts	0.158156369	-0.176018842	2.5
hn2o	0.224134231	0.481359007	2.5
hnc	0.140757699	-0.625574388	2.5
hoch3f-	0.119787346	-0.759284749	2.5
hoch4	0.124117231	-1.654285833	2.5
hocl	0.066449826	-0.32889639	2.5
hoh	0.040581623	-0.326308865	2.5
hooh	0.107512853	-0.81005108	2.5

li2	0.008827362	0.034499936	2.5
lif	0.014984765	-0.409388528	2.5
lih	0.016384451	-0.103490172	2.5
mg	0.031418131	-0.116585761	2.5
n2	0.148514254	-0.485341135	2.5
n2o	0.21491744	-0.074230644	2.5
na2	0.010639733	0.084071345	2.5
nacl	0.008425065	-0.369072359	2.5
nh	0.026123383	-0.200788602	2.5
nh2	0.049120775	-0.462978014	2.5
nh3	0.074951018	-1.069338315	2.5
no	0.120600733	-0.088265911	2.5
o2	0.104321075	0.206018895	2.5
o3	0.237334767	0.559060696	2.5
oh-	0.022555674	-0.02970047	2.5
oh	0.02496266	-0.251971797	2.5
p2	0.091496082	0.296004722	2.5
ph2	0.03485425	-0.266036342	2.5
ph3	0.04845113	-0.445113767	2.5
s2	0.04599027	0.217932807	2.5
sh	0.018690755	-0.140362644	2.5
sh3	0.041136032	-0.122407475	2.5
si	0.013338505	-0.118158203	2.5
si2	0.079984734	0.255033331	2.5
si2h6	0.096461953	-1.325164703	2.5
sih2_sing	0.051286554	-0.730936329	2.5
sih2_trip	0.032858276	-0.409903841	2.5
sih3	0.04266368	-0.638502707	2.5
sih4	0.055557927	-0.892022698	2.5
sio	0.12211427	-0.254366771	2.5
so	0.061285758	1.921983474	2.5
so2	0.130427118	1.322372332	2.5

Table 4: CBS extrapolation parameters for static correlation energies calculated using polarization consistent basis sets

	-E_infty	A	B
al	0.024078883	-0.185649498	3.2
b	0.034751335	-0.55369365	3.2
be	0.043828117	-0.807182746	3.2
beh	0.02765754	-0.771975074	3.2
c	0.019302984	-0.284459128	3.2
ch	0.043075298	-0.505157315	3.2
ch2_sing	0.06205472	-0.765165474	3.2
ch2_trip	0.038756515	-0.256614033	3.2
ch3	0.059380869	-0.372455302	3.2
ch3ch2	0.128935153	-0.629930474	3.2
ch3cl	0.087216098	-0.477950475	3.2
ch3f	0.093800637	-0.509770798	3.2
ch3fcl	0.112161277	-0.180120775	3.2
ch3sh	0.104644343	-0.474111045	3.2
ch4	0.08318303	-0.505422342	3.2
cl2	0.023274749	0.162538946	3.2
clch3cl-	0.080158853	-0.400681177	3.2
clf	0.037640865	0.087778316	3.2
clo	0.042631648	0.175288309	3.2
cn	0.151896367	-0.273424882	3.2
co	0.13177554	-0.39265239	3.2
co2	0.176294984	-0.067222451	3.2
cs	0.104574943	0.048511279	3.2
f2	0.079078282	-0.011956136	3.2
fch3cl-	0.077031475	-0.676723907	3.2
h2	0.018545613	-0.023151846	3.2
h2cch2	0.144139797	-0.595043178	3.2
h2co	0.147163489	-1.154870347	3.2
h2nnh2	0.135082998	-0.60693757	3.2
h2o	0.053734894	-0.263322901	3.2
h2s	0.034388946	-0.091001678	3.2
h3cch3	0.151166917	-0.754602814	3.2
h3coh	0.118770086	-0.592031225	3.2
hc2h4	0.148339762	-0.569416574	3.2
hcch	0.14802082	-0.548848301	3.2
hcl	0.017349087	-0.035818541	3.2
hclh	0.037399148	0.289345962	3.2
hcn	0.151465175	-0.77722357	3.2
hcn_ts	0.141785627	-0.376703261	3.2
hco	0.125812321	-0.24569144	3.2
hf	0.024434954	-0.131198476	3.2
hn2	0.129918664	-0.195929663	3.2
hn2_ts	0.158116662	-0.165302202	3.2
hn2o	0.223983023	-0.091488411	3.2
hnc	0.140766023	-0.692659978	3.2
hoch3f-	0.119717756	-0.701494566	3.2
hoch4	0.124170007	-0.619569025	3.2
hocl	0.06640562	-0.082939779	3.2
hoh	0.040600688	-0.104218287	3.2
hooh	0.107502935	-0.363791119	3.2

li2	0.008817065	0.330743392	3.2
lif	0.01498735	-0.18624612	3.2
lih	0.016385919	-0.144186746	3.2
mg	0.031232856	-0.233615137	3.2
n2	0.148527215	-0.247659394	3.2
n2o	0.214871958	-0.107018852	3.2
na2	0.010565179	0.037867926	3.2
nacl	0.008375821	0.020322854	3.2
nh	0.026111362	-0.143526475	3.2
nh2	0.049110019	-0.270466229	3.2
nh3	0.074982273	-0.424358723	3.2
no	0.120568213	-0.107651	3.2
o2	0.104271131	0.069988824	3.2
o3	0.237289444	-0.06949614	3.2
oh-	0.022486035	-0.080055139	3.2
oh	0.024948151	-0.136610435	3.2
p2	0.091325561	1.075850745	3.2
ph2	0.034790774	-0.114806731	3.2
ph3	0.048378513	-0.182618265	3.2
s2	0.045881622	0.540223146	3.2
sh	0.018660313	-0.067775521	3.2
sh3	0.041074415	-0.052012846	3.2
si	0.013274957	-0.049099559	3.2
si2	0.079833478	1.844220174	3.2
si2h6	0.096277386	-0.506810321	3.2
sih2_sing	0.051124195	-1.612552841	3.2
sih2_trip	0.032835334	-0.095011503	3.2
sih3	0.042609414	-0.19483863	3.2
sih4	0.055473823	-0.398743865	3.2
sio	0.122013817	-0.419134119	3.2
so	0.060987966	1.172677891	3.2
so2	0.130147993	0.910032664	3.2

Table 5: Numerical Hartree-Fock energies for a selection of atoms and diatomics

	Num HF	N_nu	N_mu	R_infty (Bohr)		Num HF	N_nu	N_mu	R_infty (Bohr)
H	-0.500000	169	265	35	He	-2.861680	169	265	35
Li	-7.432727	169	265	35	Na	-161.858912	169	265	35
Be	-14.573023	169	265	35	Mg	-199.614636	169	265	35
B	-24.529138	169	265	35	Al	-241.876861	169	265	35
C	-37.688694	169	265	35	Si	-288.854547	169	265	35
N	-54.400934	169	265	35	P	-340.718781	169	265	35
O	-74.812396	169	265	35	S	-397.507281	169	265	35
F	-99.411412	169	265	35	Cl	-459.483973	169	265	35
F-	-99.459454	169	265	35	Cl-	-459.576925	169	265	35
Ne	-128.547098	169	265	35	Ar	-526.817513	169	265	35
H2	-1.133624	199	355	100	Na2	-323.717017	295	499	300
HF	-100.070810	295	571	200	NaCl	-621.460596	295	523	300
LiF	-106.993356	295	559	300	FCl	-558.920172	295	571	350
F2	-198.773437	295	583	350	Cl2	-919.010838	295	553	350
N2	-108.993190	295	529	150	P2	-681.500255	295	541	300
CO	-112.790895	295	553	200	CS	-435.362393	295	559	300

Appendix 2: Supporting Information for Chapter 3

Cyclopropenium cations break the rules of attraction to form closely-bound dimers.

Computational

Methods

Potential energy curves for the dissociation and rotation of the dicationic cyclopropenium dimer were calculated in the gas phase and with partially charged counterions providing charge-balance, using Hartree-Fock (HF) Theory,[1,2] Second Order Möller-Plesset Perturbation Theory (MP2)[2,3] and Zeroth Order Symmetry-Adapted Perturbation Theory (SAPT0),[4,5] as implemented in Q-Chem (HF with CHELPG fitted atomic charges),[6] GAMESS (HF with Löwdin density partitioning, MP2)[7] and Psi4 (SAPT0).[8] The jun-cc-pVDZ basis set[9] was used throughout, as this partially augmented version of cc-pVDZ[10] provides an optimal balance between accuracy and computational cost in SAPT0 dimer calculations.[11] Local well depths for gas phase dimers were recalculated at SAPT0/juncc-pVTZ and SAPT0/aug-cc-pVDZ.

Table 1: Tris(aminoethylmethyl)cyclopropenium dimer coordinates at equilibrium (Å)

Monomer 1 coordinates			
N	7.269675	11.475827	7.759767
N	7.94432	9.209924	10.578762
N	6.587871	7.86576	7.427113
C	7.471398	9.41226	9.352562
C	7.217756	10.258966	8.286964
C	6.962495	8.90418	8.170774
C	6.530298	11.792999	6.533729
H	6.518594	11.023626	5.958871
H	6.958239	12.523392	6.081475
H	5.630012	12.038523	6.757826
C	7.93211	7.846565	11.13665
H	8.622617	7.7785	11.814625
H	8.146549	7.216484	10.431069
C	6.110452	8.099653	6.0546
H	6.254944	7.296563	5.530164
H	6.636214	8.810691	5.656015
C	6.219374	6.600267	8.064653
H	6.791992	6.442691	8.818545
H	6.317684	5.883679	7.433365
H	5.307178	6.644335	8.359795
C	8.051207	10.3296	11.518914
H	8.361919	11.110302	11.053832
H	8.672367	10.104569	12.215726
H	7.189816	10.508236	11.903127
C	6.604683	7.462274	11.749507
H	5.909485	7.557383	11.093617
H	6.420013	8.03529	12.497472
H	6.641198	6.550638	12.046842
C	7.771095	12.596373	8.568465
H	8.148133	13.265837	7.975745
H	8.485186	12.274465	9.140887
C	6.709508	13.249784	9.431483
H	6.030861	13.631651	8.869697
H	7.110482	13.941281	9.963388
H	6.314655	12.590469	10.007234
C	4.644443	8.474126	5.984041
H	4.404058	8.655833	5.072226
H	4.487958	9.256423	6.518302
H	4.11285	7.748122	6.316939
Monomer 2 coordinates			
N	10.18389	6.983513	8.479151
N	9.509245	9.249416	5.660156
N	10.865695	10.59358	8.811806
C	9.982168	9.04708	6.886357
C	10.23581	8.200374	7.951955
C	10.491071	9.55516	8.068144
C	10.923267	6.666341	9.70519
H	10.934972	7.435715	10.280047
H	10.495327	5.935948	10.157444
H	11.823554	6.420817	9.481093

C	9.521455	10.612775	5.102268
H	8.830949	10.68084	4.424293
H	9.307017	11.242856	5.807849
C	11.343113	10.359687	10.184319
H	11.198622	11.162777	10.708755
H	10.817351	9.648649	10.582903
C	11.234192	11.859073	8.174266
H	10.661574	12.01665	7.420374
H	11.135882	12.575661	8.805554
H	12.146388	11.815005	7.879123
C	9.402358	8.12974	4.720004
H	9.091647	7.349039	5.185087
H	8.781198	8.354772	4.023192
H	10.26375	7.951104	4.335791
C	10.848882	10.997066	4.489411
H	11.544081	10.901957	5.145301
H	11.033552	10.42405	3.741447
H	10.812368	11.908702	4.192077
C	9.68247	5.862967	7.670453
H	9.305433	5.193503	8.263174
H	8.968379	6.184875	7.098031
C	10.744057	5.209556	6.807436
H	11.422704	4.827689	7.369221
H	10.343083	4.518059	6.27553
H	11.138911	5.868871	6.231685
C	12.809123	9.985214	10.254877
H	13.049507	9.803508	11.166692
H	12.965607	9.202918	9.720617
H	13.340715	10.711218	9.921979

Table 2: Coordinates of surrounding counterions at equilibrium (Å)

Monomer 1 associated				Monomer 2 associated			
Cl	6.581766	9.592796	2.832311	Cl	7.420913	4.486778	5.287148
Cl	10.032653	13.972562	10.95177	Cl	10.8718	8.866544	13.406608
Cl	8.332961	4.272754	10.95177	Cl	9.120605	14.186586	5.287148

Table 3: Atomic partial charges from atom-centred density partitioning (Löwdin) and electrostatic potential fitting (CHELPG) of the HF/jun-cc-pVDZ electron density

dimer, HF/jun-cc-pVDZ			monomer, HF/jun-cc-pVDZ		
	Lowdin	CHELPG		Lowdin	CHELPG
N	-0.123263	-0.270437	N	-0.110259	-0.359879
N	-0.128064	-0.357782	N	-0.114931	-0.400638
N	-0.127437	-0.353428	N	-0.114264	-0.408145
C	-0.001622	0.157362	C	0.008984	0.175728
C	-0.000266	0.075736	C	0.010223	0.141602
C	0.006377	0.150163	C	0.016649	0.178275
C	-0.114226	-0.052665	C	-0.109917	-0.055288
H	0.115944	0.07449	H	0.111741	0.105536
H	0.119373	0.082575	H	0.116328	0.092386
H	0.11022	0.089793	H	0.104098	0.082277
C	-0.104779	0.478095	C	-0.089932	0.364776
H	0.114387	-0.035246	H	0.112203	-0.008548
H	0.115658	-0.092098	H	0.110486	-0.039876
C	-0.104605	0.442972	C	-0.091344	0.369552
H	0.113745	-0.026632	H	0.111637	-0.007615
H	0.116454	-0.067604	H	0.111687	-0.040407
C	-0.116741	-0.04692	C	-0.112825	-0.016065
H	0.117237	0.068243	H	0.113097	0.095982
H	0.117852	0.085672	H	0.115044	0.084678
H	0.109542	0.094258	H	0.103467	0.073215
C	-0.119886	-0.064158	C	-0.116945	-0.028248
H	0.116911	0.076317	H	0.112235	0.098133
H	0.119509	0.088418	H	0.116402	0.089356
H	0.112191	0.099773	H	0.106235	0.076934
C	-0.204522	-0.160455	C	-0.21655	-0.004961
H	0.103527	0.044489	H	0.100917	-0.002073
H	0.10644	0.054911	H	0.102501	0.017677
H	0.098784	0.047592	H	0.096045	-0.001208
C	-0.107136	0.428255	C	-0.093009	0.363153
H	0.11531	-0.029654	H	0.11305	-0.010104
H	0.114419	-0.070588	H	0.109914	-0.040117
C	-0.21188	-0.167646	C	-0.223375	-0.02691
H	0.106055	0.057923	H	0.10214	0.024838
H	0.105888	0.053724	H	0.103064	0.00524
H	0.103693	0.045935	H	0.101338	0.003379
C	-0.205222	-0.169682	C	-0.217205	-0.026976
H	0.099528	0.049829	H	0.096605	0.003235
H	0.103509	0.049392	H	0.101103	0.00597
H	0.107095	0.060956	H	0.103362	0.025137

Table 4: HF/jun-cc-pVDZ, SAPT0/jun-cc-pVDZ and BSSE-corrected MP2/jun-ccpVDZ energies (Eh) as a function of displacement along intermolecular C3-C3 centroid vector.

$\Delta r_{C_3-C_3}$ (Å)	HF	SAPT0	MP2
-0.5	-1262.2368841	0.089397055	-1266.537054
-0.4	-1262.2551331	0.07919667	-1266.548901
-0.3	-1262.2683993	0.072647887	-1266.556706
-0.2	-1262.2780305	0.068630891	-1266.561677
-0.1	-1262.2852202	0.066332051	-1266.564751
0	-1262.2905177	0.065167634	-1266.566506
0.1	-1262.2942701	0.064724894	-1266.567359
0.2	-1262.2970610	0.064717001	-1266.567672
0.3	-1262.2991733	0.064948611	-1266.567664
0.4	-1262.3008102	0.065289911	-1266.567485
0.5	-1262.3021179	0.065657349	-1266.567233
0.6	-1262.3032008	0.065999347	-1266.566971
0.7	-1262.3041322	0.066286072	-1266.566737
0.8	-1262.3049630	0.066502148	-1266.566554
0.9	-1262.3057271	0.066641392	-1266.566433
1	-1262.3064463	0.066703455	-1266.566379
.	.	.	.
.	.	.	.
.	.	.	.
2.0	-1262.312644	0.06687*	-1266.568746
.	.	.	.
.	.	.	.
.	.	.	.
∞	-1262.377878	0.0	-1266.632645

*The SAPT0 energy at 2.0 Å displacement has been extrapolated using a second order expansion in cubed inverse bond lengths. Note also that SAPT0 directly computes the interaction energy relative to infinite separation of the monomer units, whereas HF and MP2 energies are absolute.

Table 5: SAPT0/jun-cc-pVDZ interaction energy components (mEh) at and near the crystal structure geometry of 1. Slightly separated (positively displaced) geometries cover the region in which the gas phase minimum energy is found. The dimer is only meta-stable in the absence of the surrounding counterions, as indicated by the overall positive total energy here.

	$\Delta r_{C_3-C_3}$			
	0.0	0.1	0.2	0.3
Electrostatics	80.1	81.0	81.3	81.3
Exchange	19.8	14.1	10.0	7.1
Induction	-6.1	-5.4	-4.9	-4.4
Dispersion	-28.7	-24.9	-21.7	-19.0
Total	65.1	64.7	64.7	64.9

Table 6: As above, energies in kJ mol^{-1}

	$\Delta r_{C_3-C_3}$			
	0.0	0.1	0.2	0.3
Electrostatics	210.4	212.6	213.5	213.3
Exchange	52.0	36.9	26.2	18.5
Induction	-16.0	-14.2	-12.8	-11.6
Dispersion	-75.3	-65.4	-57.0	-49.8
Total	171.1	169.9	169.9	170.5

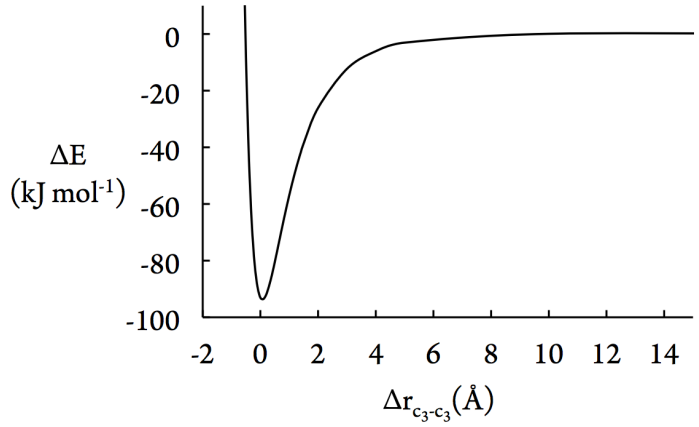


Figure 1: BSSE-corrected MP2/jun-cc-pVDZ interaction potential for dimeric complex of compound 1, each with three associated partially-charged Cl^- ions (Zeff = 17.666). The predicted minimum lies within 0.1 Å of the observed C_3-C_3 centroid distance, validating the reliability of our partial charge model for the crystalline environment.

Table 7: BSSE-corrected MP2 energies (E_h) as a function of displacement along intermolecular C_3-C_3 centroid vector for rotated cyclopropenium units ($\Delta\theta = 60^\circ$).

$\Delta r_{C_3-C_3}$ (Å)	MP2
-0.4	-1266.518556
-0.3	-1266.521300
-0.2	-1266.523911
-0.1	-1266.537509
0	-1266.546803
0.1	-1266.553093
0.2	-1266.557315
0.3	-1266.560107
0.4	-1266.561941
0.5	-1266.563141
0.6	-1266.563929
0.7	-1266.564458
0.8	-1266.564832
0.9	-1266.565119
1	-1266.565366

Table 8: BSSE-corrected MP2/jun-cc-pVDZ energies (E_h) as a function of displacement along intermolecular C_3-C_3 centroid vector for varying sizes of amino groups.

$\Delta r_{C_3-C_3}$ (Å)	NEt2	NEtMe	NMe2	NH2
-0.5	-1501.485723	-1266.537054	-1032.213379	-562.0678006
-0.4	-1501.495124	-1266.548901	-1032.218252	-562.0751602
-0.3	-1501.501207	-1266.556706	-1032.221219	-562.0806179
-0.2	-1501.504960	-1266.561677	-1032.222898	-562.0846816
-0.1	-1501.507141	-1266.564751	-1032.223731	-562.0877348
0	-1501.508277	-1266.566506	-1032.224032	-562.0900662
0.1	-1501.508755	-1266.567359	-1032.224020	-562.0918923
0.2	-1501.508806	-1266.567672	-1032.223844	-562.0933720
0.3	-1501.508624	-1266.567664	-1032.223601	-562.0946182
0.4	-1501.508328	-1266.567485	-1032.223355	-562.0957156
0.5	-1501.508017	-1266.567233	-1032.223144	-562.0967201
0.6	-1501.507717	-1266.566971	-1032.222989	-562.0976713
0.7	-1501.507434	-1266.566737	-1032.222901	-562.0985922
0.8	-1501.507210	-1266.566554	-1032.222883	-562.0994985
0.9	-1501.507053	-1266.566433	-1032.222934	-562.1003998
1	-1501.506965	-1266.566379	-1032.223050	-562.1013023
1.1	-1501.506944	-1266.566391	-1032.223227	-562.1022077
1.2	-1501.506988	-1266.566467	-1032.223458	-562.1031149
1.3	-1501.507090	-1266.566603	-1032.223737	-562.1040241
1.4	-1501.507246	-1266.566793	-1032.224058	-562.1049341
1.5	-1501.507449	-1266.567030	-1032.224416	-562.1058437
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
2	-1501.508987	-1266.568746	-1032.226584	-562.1103209
Well depth:	0.001862316	0.001293008	0.001149177	no well present

Table 9: SAPT0/jun-cc-pVDZ local well positions for dissociating gas phase TAC dicationic dimers, in Å relative to experimentally determined thermally averaged distance between C_3 centroids in complex **1**

Amino groups	Position of local minimum	Position of local maximum
NEt2	0.08	1.06
NEtMe	0.15	1.02
NMe2	-0.03	0.77

Table 10: SAPT0 and BSSE-corrected MP2 local well depths for dissociating TAC dimers, in kJ mol^{-1}, calculated at geometries specified above. The SAPT0 results are considered more reliable as they are inherently BSSE-free, so do not suffer from BSSE correction errors, particularly for the larger basis calculations.

Amino groups	MP2/jun-cc-pVDZ	MP2/aug-cc-pVDZ	MP2/jun-cc-pVTZ
NEt2	4.9	8.9	9.7
NEtMe	3.5	9.3	9.6
NMe2	3.1	9.6	9.8
Amino groups	SAPT0/jun-cc-pVDZ	SAPT0/aug-cc-pVDZ	SAPT0/jun-cc-pVTZ
NEt2	6.7	14.1	14.8
NEtMe	5.3	12.2	12.9
NMe2	4.3	11.3	12.1

Molecular orbital analysis

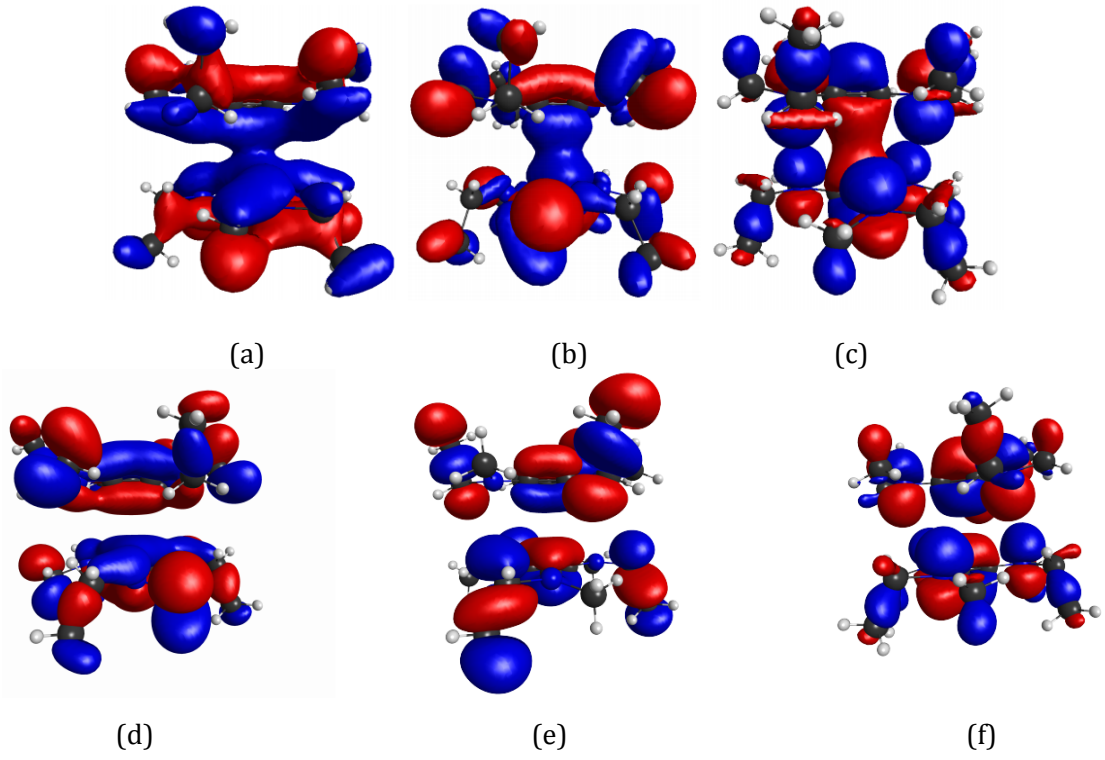


Figure 2: (a)-(c) Hartree-Fock π-bonding orbitals and (d)-(f) π^*-antibonding orbitals of the tris(aminoethylmethyl)cyclopropenium dimer at equilibrium geometry. All depicted orbitals are doubly occupied in the HF calculation.

Synthesis

General

All operations were performed using standard Schlenk techniques with a dinitrogen atmosphere to reduce exposure to water. ^{1}H-, ^{13}C{^{1}H}-NMR spectra were collected on an Agilent DD2-400MR operating at 400 and 100 MHz, respectively, in $CDCl_3$, referenced to residual solvent peaks. Electrospray mass spectrometry was carried out on a Micromass LCT, with samples dissolved in acetonitrile. Microanalysis was performed by Campbell Microanalytical Laboratory, Dunedin. Pentachlorocyclopropane was obtained from Acros, while ethylmethylamine was prepared by methods described by Lucier and Wawzonek.[12,13]

Synthesis of tris(ethylmethylamino)cyclopropenium chloride

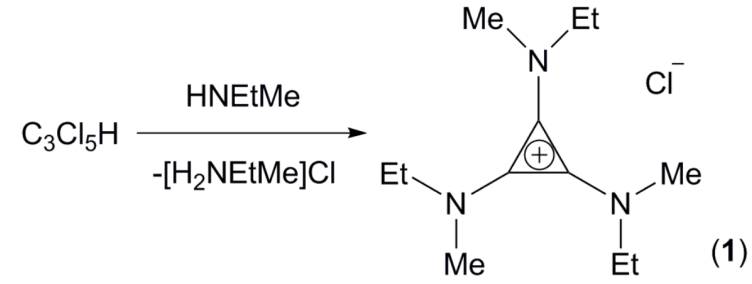
A 3:1 molar ratio of ethylmethylamine and dimethylethylamine (20.0 g, ca. 0.2 mol NEtMeH) in dichloromethane (150 mL) was cooled to 0 °C. C3Cl5H (0.024 moles, 5.15 g) was added drop-wise and the solution was stirred overnight at ambient temperature and then heated to reflux for 3.5 hours. The dichloromethane was removed in vacuo and then dissolved in water (50 mL). The pH was increased to 11.0 using 12.5 M NaOH(aq) before washing with diethyl ether (3 x 20 mL). The aqueous solution was neutralized with 6M HCl(aq) before extracting the product with dichloromethane (3 x 20 mL). The dichloromethane was removed in vacuo to give an orange liquid. Water (15 mL) was added to the mixture. The pH of the aqueous solution was adjusted to 1 using 6 M HCl(aq) and the product was extracted with chloroform (3 x 20 mL). The chloroform was removed in vacuo to give the product as an orange liquid $[C_3(NEtMe)_3]Cl$ (1.82 g, 30.86%). ESI MS: m/z 210.19 (M+); ^{1}H NMR ($CDCl_3$, 400 MHz): δ 3.47 (q, $^{3}J_{HH}$ = 7.4 Hz, 6H, NCH_2CH_3), 3.19 (s, 9H, NCH_3), 1.29 (t, $^{3}J_{HH}$ = 7.4 Hz, 9H, NCH_2CH_3). ^{13}C{1H} NMR ($CDCl_3$, 400 MHz): δ 116.8 (C3), 50.23

(NCH_2CH_3), 39.42 (NCH_3), 13.10 (NCH_2CH_3). Found: C, 42.99; H, 7.95; N, 11.59%. Calc. for $C_{12}H_{24}N_3Cl_5 \cdot 5H_2O$: C, 42.79; H, 10.20; N, 12.47%).

Synthesis of tris(diethylamino)cyclopropenium iodide

$[C_3(NEt_2)_3]Cl$ (1.15 g, 4 mmol) was heated to reflux with ethyl iodide (6.43 mL, 80 mmol) for 20 h in an inert atmosphere. Unreacted ethyl iodide was distilled out and the residue was dissolved in CH_2Cl_2 (20 mL) and washed with water (3 Œ 10 mL). CH_2Cl_2 was removed in vacuo to give orange crystals (1.2 g, 79%). 1H NMR (500 MHz, $CDCl_3$): δ 3.44 (q, $^3J_{HH}$ = 7 Hz, 12H, NCH_2), 1.30 (t, $^3J_{HH}$ = 7 Hz, 18H, NCH_2CH_3). $^{13}C\{1H\}$ NMR (126 MHz, CD_3CN): δ 116.13 (ring C), 47.17 (CH_2), 14.42 (CH_3). EI MS: m/z 252.2422 (M+), calc'd 252.2434. Anal. calcd for $C_{15}H_{30}N_3I$: C, 47.50; H, 7.97; N, 11.08; I, 33.45. Found: C, 48.44; H, 7.97; N, 11.07; I, 32.35. Iodide content was determined by ion chromatography; chloride content was found to be 14 ppm.

Reaction scheme



$[C_3(NEt_2)_3]Cl + EtI \longrightarrow [C_3(NEt_2)_3]I + EtCl$

(**2**)

Figure 3: Synthesis of salts **1** and **2**

Crystallographic Data

Single crystals of $C_{12}H_{24}ClN_3$ and $C_{15}H_{30}IN_3$ formed in the neat liquid. A suitable crystal was selected and mounted on a nylon loop in perfluoronated oil on a

SuperNova, Dual, Cu at zero, Atlas diffractometer. The crystal was kept at 120.02(10) K during data collection. Using Olex2,[14] the structure was solved with the olex2.solve structure solution program using Charge Flipping and refined with the ShelXL[15] refinement package using Least Squares minimisation. Crystal data and structure refinement details are given in Table S1. Structural details are available through the Cambridge Crystallographic Data Centre (CCDC codes 1062303 and 1062302).

Table 11: Crystal data and structure refinement for $[C_3(NEtMe)_3]Cl$ and $[C_3(NEt_2)_3]I$

Identification code	CHAM1ma	CHAM2Ra
Empirical formula	$C_{12}H_{24}ClN_3$	$C_{15}H_{30}IN_3$
Formula weight	245.79	379.32
Temperature/K	120.02(10)	120.01(10)
Crystal system	triclinic	monoclinic
Space group	P-1	C2/c
a/Å	8.8521(3)	18.4506(4)
b/Å	9.8476(3)	10.6276(3)
c/Å	9.8497(5)	19.3662(5)
α/°	60.102(4)	90
β/°	69.491(4)	104.431(3)
γ/°	80.061(3)	90
Volume/Å^3	697.16(6)	3677.61(16)
Z	2	8
ρ_{calc} mg/mm^3	1.171	1.370
m/mm^{-1}	2.253	13.615
F(000)	268.0	1552.0
Crystal size/mm^3	0.171 × 0.1219 × 0.0858	0.1967 × 0.1095 × 0.0949
Radiation	CuKα (λ = 1.54184)	CuKα (λ = 1.54184)
2θ range for data collection	10.362 to 149.49°	9.432 to 154.2°
Index ranges	-10 ≤ h ≤ 11, -12 ≤ k ≤ 12, -11 ≤ l ≤ 12	-19 ≤ h ≤ 23, -13 ≤ k ≤ 13, -21 ≤ l ≤ 24
Reflections collected	11244	14985
Independent reflections	2783 [R_{int} = 0.0320, R_{sigma} = 0.0226]	3870 [R_{int} = 0.0313, R_{sigma} = 0.0257]
Data/restraints/parameters	2783/0/151	3870/0/178
Goodness-of-fit on F^2	1.050	1.026
Final R indexes [I ≥ 2σ (I)]	R_1 = 0.0297, wR_2 = 0.0784	R_1 = 0.0234, wR_2 = 0.0613
Final R indexes [all data]	R_1 = 0.0337, wR_2 = 0.0810	R_1 = 0.0263, wR_2 = 0.0634
Largest diff. peak/hole / e Å^{-3}	0.38/-0.19	0.52/-0.80

Illustrated crystal structure of 2

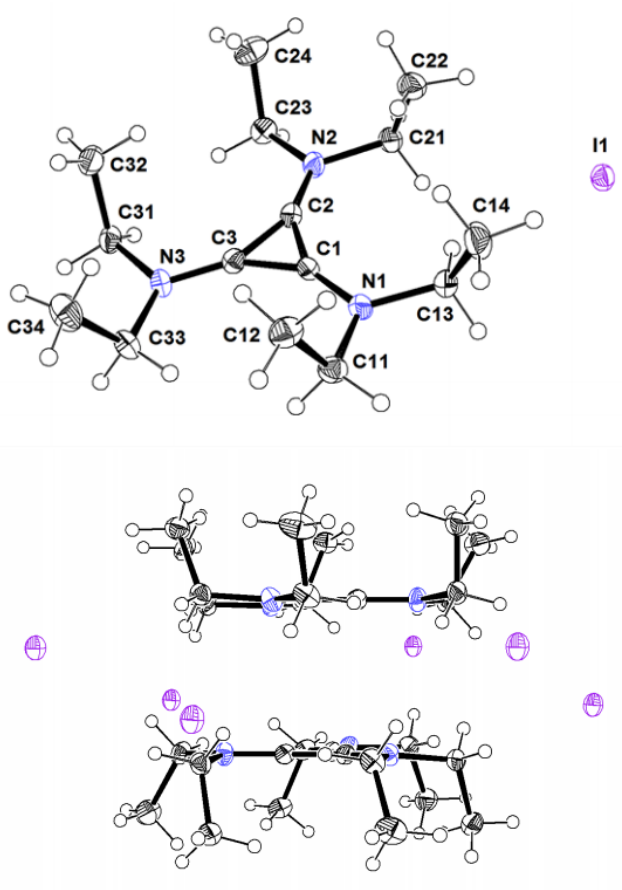


Figure 4: (a) The cation of **2** with the atomic numbering scheme, (b) side-on view of the dicationic dimer unit with surrounding counterions. Atoms are represented as 40% probability thermal ellipsoids.

References

(1) Roothaan, C. C. *J. Rev. Mod. Phys.* **1951**, *23*, 69 – 89.

(2) Szabo, A.; Ostlund, N. S. Modern Quantum Chemistry; McGraw-Hill: New York, **1989**.

(3) Pople, J. A.; Binkley, J. S.; Seeger, R. Int. J. Quantum Chem. **1976**, *S10*, 1 – 19.

(4) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887 – 1930.

(5) Szalewicz, K. *WIREs Comput. Mol. Sci.* **2012**, *2*, 254 – 272.

(6) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kús, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock III, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S.

H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio Jr., R. A.; Dop, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Pieniazek, P. A.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sergueev, N.; Sharada, S. M.; Sharmaa, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Vanovschi, V.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhou, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; Goddard III, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer III, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xua, X.; AspuruGuzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Van Voorhis, T.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; HeadGordon, *M. Mol. Phys.* **2015**, *113*, 184–215.

(7) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1363 – 1363.

(8) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. *WIREs Comput. Mol. Sci.* **1993**, *14*, 556 – 1363.

(9) Papajak, E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2011**, *7*, 10 – 18.

(10) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007 – 1023.

(11) Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. *J. Chem. Phys.* **2014**, *140*, 094106.

(12) Lucier, J. J.; Harris, A. D.; Korosec, P. S. N-Methylbutylamine. *Organic Syntheses* **1964**, 72–72.

(13) Wawzonek, S.; McKillip, W.; Peterson, C. N-Methylethylamine. *Organic Syntheses* **1964**, 75–75.

(14) Dolomanov, O. V.; Bourhis, L. J.; Gildea, R. J.; Howard, J. A. K.; Puschmann, H. *J. Appl. Crystallogr.* **2009**, *42*, 339–341.

(15) Sheldrick, G. M. *Acta Cryst.* **2008**, *A64*, 112 – 122

Appendix 3: Supporting Information for Chapter 4

The geometry of zinc phthalocyanine was optimized using B3LYP/6-31G*. D_{4h} symmetry was subsequently enforced. Comparison to crystallographic data determined this geometry was within the range of existing crystal geometries for all metallophthalocyanines. We therefore used this geometry for calculations on all metallophthalocyanines, changing only the central metal.

Metal	0.00000	0.00000	0.00000
N	1.96332	0.00000	0.00000
C	2.76952	1.11759	0.00000
N	2.38838	2.38838	0.00000
C	4.16597	0.70401	0.00000
C	5.36092	1.42570	0.00000
H	5.35189	2.51120	0.00000
C	6.55323	0.70437	0.00000
H	7.50160	1.23482	0.00000
C	2.76952	-1.11759	0.00000
C	4.16597	-0.70401	0.00000
C	5.36092	-1.42570	0.00000
H	5.35189	-2.51120	0.00000
C	6.55323	-0.70437	0.00000
H	7.50160	-1.23482	0.00000
N	0.00000	1.96332	0.00000
C	-1.11759	2.76952	0.00000
N	-2.38838	2.38838	0.00000
C	-0.70401	4.16597	0.00000
C	-1.42570	5.36092	0.00000
H	-2.51120	5.35189	0.00000
C	-0.70437	6.55323	0.00000
H	-1.23482	7.50160	0.00000
C	1.11759	2.76952	0.00000
C	0.70401	4.16597	0.00000
C	1.42570	5.36092	0.00000
H	2.51120	5.35189	0.00000
C	0.70437	6.55323	0.00000
H	1.23482	7.50160	0.00000
N	-1.96332	0.00000	0.00000
C	-2.76952	-1.11759	0.00000
N	-2.38838	-2.38838	0.00000
C	-4.16597	-0.70401	0.00000
C	-5.36092	-1.42570	0.00000

H	-5.35189	-2.51120	0.00000
C	-6.55323	-0.70437	0.00000
H	-7.50160	-1.23482	0.00000
C	-2.76952	1.11759	0.00000
C	-4.16597	0.70401	0.00000
C	-5.36092	1.42570	0.00000
H	-5.35189	2.51120	0.00000
C	-6.55323	0.70437	0.00000
H	-7.50160	1.23482	0.00000
N	0.00000	-1.96332	0.00000
C	1.11759	-2.76952	0.00000
N	2.38838	-2.38838	0.00000
C	0.70401	-4.16597	0.00000
C	1.42570	-5.36092	0.00000
H	2.51120	-5.35189	0.00000
C	0.70437	-6.55323	0.00000
H	1.23482	-7.50160	0.00000
C	-1.11759	-2.76952	0.00000
C	-0.70401	-4.16597	0.00000
C	-1.42570	-5.36092	0.00000
H	-2.51120	-5.35189	0.00000
C	-0.70437	-6.55323	0.00000
H	-1.23482	-7.50160	0.00000

Table 1: Abolute energies for the calculation reported in chapter 4, table 5, EOM-CCSD/3-21G energies and dimensionless oscillator strengths (*f*), of symmetry-allowed B_{1u} and B_{2u}/B_{3u} states.

Reported relative energies		Absolute energies
Energy (eV)	*f*	(hartrees)
0.00		-3420.548547
2.42	0.41	-3420.459543
4.34	1.07	-3420.389223
4.52	0.22	-3420.382595
4.57 (B_{1u})	0.01	-3420.371595
4.82	0.40	-3420.362772
5.06	0.15	-3420.353893
5.30	0.10	-3420.339017
5.70	0.18	-3420.321990
6.16	0.11	-3420.293081
6.85 (B_{1u})	0.01	-3420.291355
6.95	0.05	-3420.548547
7.00	0.68	-3420.459543

Table 2: Energies (eigenvalues) of the molecular orbitals in chapter 4, Figure 5. ZnPc HF/6-31G* molecular orbitals near the HOMO and LUMO orbitals. Orbital number in the HF/6-31G* energy ordering is given, along with D_{2h} symmetry labels. The absolute RHF energy is -3434.335350.

Orbital number	Eigenvalues (hartree)	Orbital symmetry (d_{2h})	Eigenvalues (eV)
132-133	-0.4316	30b2u/b3u	-11.74
136	-0.412	25b1g	-11.21
137	-0.4005	37ag	-10.9
138	-0.3619	2au	-9.85
139	-0.3528	7b1u	-9.6
140	-0.339	3au	-9.22
141-142	-0.3372	5b2g/b3g	-9.18
143-44	-0.3331	6b2g/b3g	-9.06
145	-0.33	8b1u	-8.98
146	-0.3238	9b1u	-8.81
147	-0.1852	4au	-5.04
148-9	-0.0034	7b2g/b3g	-0.09
150	0.0937	5au	2.55
151	0.1053	10b1u	2.87
152	0.1059	11b1u	2.88
153-4	0.1209	8b2g/b3g	3.29
158-9	0.187	9b2g/b3g	5.09

Table 3: Absolute energies for calculations reported in Chapter 4, Table 7: EOM-CCSD/3-21G energies (in eV), with some additional orbitals frozen, benchmarked against standard calculations where only the chemical core was frozen (shaded).

EOM-CCSD/3-21G energies (eV) with various frozen orbitals

Orbitals:					
#active occ	98	98	75	32	18
#active virt	167	100	167	167	37
#frozen occ	49 (core)	49 (core)	72	115	129
#frozen virt	105	172	105	105	235
	Ground State energies				
	-3419.419985	-3418.436339	-3418.764751	-3417.372284	-3435.313355
	B_{2u}/B_{3u} states				
	-3419.328374	-3418.344626	-3418.673218	-3417.276848	-3435.219485
	-3419.261605	-3418.274871	-3418.604859	-3417.203011	-3435.143117
	-3419.253395	-3418.268387	-3418.597624	-3417.198198	-3435.131025
	-3419.241784	-3418.257186	-3418.586239	-3417.189460	-3435.126745
	-3419.232802	-3418.240272	-3418.577786	-3417.181020	-3435.122084
	B_{1u} states				
	-3419.250907	-3418.266749	-3418.591162	-3417.179636	-3435.10389

Table 4: Absolute energies for the calculations reported in chapter 4, table 8. CCSD excitation energies (eV) in various basis sets.

	3-21G	3-21G*	6-31G*	cc-pVDZ
Orbitals:				
#active occ	98	98	98	50
#active virt	272	307	167	107
#frozen occ	49	49	49	97
#frozen virt	0	172	312	475
		Ground State		
	-3420.548547	-3421.881686	-3436.963678	-3435.342851
		Computational - B_{2u}/B_{3u} states		
	-3420.459543	-3421.795147	-3436.876016	-3435.252210
	-3420.389223	-3421.723534	-3436.806074	-3435.178208
	-3420.382595	-3421.716439	-3436.800202	
	-3420.362772	-3421.706527	-3436.790649	
	-3420.353893	-3421.697539	-3436.782252	
	-3420.339017		-3436.779304	
	-3420.321990		-3436.758402	
		Computational - B_{1u} states		
	-3420.371595		-3436.785202	

Table 5: Absolute energies for the calculation results in chapter 4, table 9. Effects on excited state energies of adding diffuse functions to the centre of mass (CM).

	CM-augmented CCSD/3-21G
Orbitals:	
#active occ	98
#active virt	167
#frozen occ	49
#frozen virt	123
	Ground state
	-3419.278317
	B_{2u}/B_{3u} states
	-3419.188185
	-3419.120838
	-3419.114216
	-3419.101923
	-3419.092947
	-3419.083213
	-3419.067995
	-3419.051954

Table 6: Absolute energies for the calculation results in chapter 4, Table 10. Various EOM-CCSD(T)/3-21G calculations using different algorithms

	QChem CCSD(fT)	GAMESS CR-EOML (2,3),D	NWChem CR-EOM-CCSD(T)	GAMESS CR-EOML DEL(IID)
Orbitals:				
#active occ	98	50	98	50
#active virt	167	100	167	100
#frozen occ	49	97	49	97
#frozen virt	105	172	105	172
		Ground State		
	-3419.535080	-3417.571125	-3419.419989	-3417.553581
		B_{2u}/B_{3u} states		
	-3419.350327	-3417.455005	-3419.337617	-3417.455652
		-3417.382868	-3419.268057	-3417.387776
		-3417.377512	-3419.260958	-3417.382690
		-3417.366799	-3419.249083	-3417.371246
		-3417.355155	-3419.240948	-3417.368256
				-3417.358967
				-3417.340375
				-3417.324323
				-3417.306071
				-3417.296072
		B_{1u} states		
	-3419.267774		-3419.255082	-3417.888768*
				-3417.826402*

* Only 72 occupied orbitals were frozen in this calculation, as Table 6 indicated B_{1u} states energies are sensitive to having too many frozen occupied orbitals.

Table 7: Absolute energies for the calculation results in chapter 4, table 11. EOM-CCSD results for a minimal basis set (MBS)

EOM-CCSD/MBS	
Relative energies (eV)	Absolute energies (eV)
2.77	-3432.758333
4.70	-3432.687529
5.18	-3432.669995
4.57*	-3432.692349
5.40	-3432.661720
5.85	-3432.645243
6.36	-3432.626341
6.70	-3432.613984
6.51*	-3432.620916
6.63	-3432.616596

* States of B_{1u} symmetry

Appendix 4: Supporting Information for Chapter 5

See Appendix 3 for the geometry used for calculations on the metallophthalocyanines.

Table 1: Absolute energies for the calculation results in chapter 5, table 1. The CASSCF and MRMP2 results for the energies of the Q transition in different bases. Active space consists of the HOMO, HOMO-1, LUMO, LUMO+1, and LUMO+2 orbitals.

Method:	CASSCF	CASSCF	CASSCF	CASSCF
Basis set:	3-21G	3-21G*	cc-pVDZ	cc-pVTZ
Ground state:	-3416.63185	-3417.506202	-3434.829706	-3435.190429
Excited state:	-3416.517319	-3417.384401	-3434.719975	-3435.078273
Method:	MRMP2	MRMP2	MRMP2	
Basis set:	3-21G	3-21G*	cc-pVDZ	
Ground state:	-3420.489281	-3423.050289	-3440.697739	
Excited state:	-3420.415739	-3422.980284	-3440.632712	

Table 2: Absolute energies for the calculation results in chapter 5, table 2. CASSCF/cc-pVDZ calculation was performed using initial orbitals from a prior HF calculation with the 142 lowest energy orbitals frozen, and the next 12 orbitals comprising the active space. The first 18 states were equally weighted in the calculation

Ground state energy: -3434.837152

Calculated relative energies (eVs):	Absolute energies
3.02	-3434.726059
3.54	-3434.70698
4.58	-3434.669014
4.78	-3434.661613
5.74	-3434.626337
6.03	-3434.615414
6.13	-3434.612018
6.31	-3434.605219
6.40	-3434.601824
6.68	-3434.59175

Table 3: Absolute energies for the calculations in chapter 5 table 3. Results from CASCI calculations using different active spaces and basis sets. Orbitals used were generated from HF calculations. The active orbitals are the highest energy occupied orbitals and lowest energy virtual orbitals.

Active orbitals			
Occupied:	18	18	38
Virtual:	3	3	3
Basis:	3-21G	cc-pVDZ	3-21G
	Ground state energies:		
	-3416.646012	-3434.848227	-3416.651910
	Excited state energies:		
	-3416.536983	-3434.743077	-3416.544946
	-3416.457719	-3434.662774	-3416.466662
	-3416.443336	-3434.644293	-3416.452904
	-3416.438530	-3434.641735	-3416.447781
	-3416.423876	-3434.624627	-3416.432654

Active orbitals			
Occupied:	31	9	11
Virtual:	3	6	6
Basis:	cc-pVDZ	cc-pVDZ	3-21G
	Ground state energies:		
	---	-3434.856003	-3416.670216
	Excited state energies:		
	-3434.752486	-3434.750108	-3416.559070
	-3434.673736	-3434.673142	-3416.482323
	-3434.654802	-3434.659159	-3416.466514
	-3434.653998	-3434.654644	-3416.462041
	-3434.634793	-3434.641761	---

Table 4: Absolute energies for the calculation results reported in chapter 5, table 4. The active space for these calculations comprised 7 active orbitals, and 8 electrons. The HF orbitals used in the active space were: 119 $1a_u$, 139 $7b_{1u}$, 146 $9b_{1u}$, 147 $4a_u$ (HOMO), 148-9 $7b_{2g}/b_{3g}$ (LUMO), 150 $5a_u$. The GAMESS program provides an energy denominator offset (ε) which can be set to prevent division-by-near-zero due to near-degeneracies in energy occurring in the perturbation theory energy correction. Results can also be affected by whether the perturbation theory energy correction is applied to each state independently (called MRMP2) or to multiple states together (called MCQDPT2).

Excited state #		Ground	First	Second
CASSCF		-3434.831325	-3434.720293	-3434.603885
MCQDPT2	ε = 0.0	-3440.699408	-3440.674625	-3440.563728
	ε = 0.002	-3440.698633	-3440.667859	-3440.560487
	ε = 0.02	-3440.691799	-3440.652176	-3440.543647
	ε = 0.05	-3440.680941	-3440.634260	-3440.523966

Appendix 5: Supporting Information for Chapter 6

See Appendix 3 for the geometry used for calculations on the metallophthalocyanines.

Table 1: Absolute energies of the CASSCF/cc-pVDZ and MRMP2/cc-pVDZ calculations on CoPc reported in chapter 6, table 1.

Active Space		Calculations		
#occ	Metal	3	4	3
	Ligand	1	2	1
#virt	Ligand	2	3	4
	2nd set of metal	0	0	0
Ground CASSCF		-3038.401499	-3038.414724	-3038.401658
Ground MRMP2		-3044.085092	-3044.099879	-3044.085978
First excited state CASSCF		-3038.379607	-3038.392633	-3038.379746
First excited state MRMP2		-3044.071931	-3044.086945	-3044.072822
Active Space		Calculations		
#occ	Metal	3	3	4
	Ligand	6	1	6
#virt	Ligand	2	2	4
	2nd set of metal	0	3	3*
Ground CASSCF		-3038.412336	-3038.459160	-3038.420585
Ground MRMP2		-3044.087018	-3044.142628	-3044.088678
First excited state CASSCF		-3038.390447	-3038.442237	-3038.399571*
First excited state MRMP2		-3044.074951	-3044.132455	-3044.076924*

* This calculation was done at the CISDTQ level rather than CASSCF. i.e. orbitals were frozen and a maximum of 4 excitations was imposed, in order to reduce the computational memory requirements of the algorithm for the large active space. The occupied d_{xy} orbital was included in the active space, but its virtual counterpart was not.

Table 2: Absolute energies of low-lying FePc states reported in chapter 6, table 2. CASSCF/cc-pVDZ and MRMP2/cc-pVDZ calculations were performed in active spaces containing the specified number of highest-occupied and lowest-unoccupied ligand orbitals, and the specified number of highest-occupied metal orbitals and sometimes unoccupied duplicates of the metal orbital from the virtual orbitals (the cc-pVDZ basis provides duplicate copies of each AO). States are characterized by their D_{4h} symmetry label and by their singly occupied molecular orbitals (SOMOs).

Active Space		Calculations		
#occ	Metal	4	4	3
	Ligand	1	1	1
#virt	Ligand	2	2	2
	2nd set of metal	0	0	3
#States weighted:		3	4	3
State and singly occupied metal MOs				
$^3A_{2g}$	d_{xz} and d_{yz}			
		-2919.422167	-2919.421413	-2919.461398
		-2925.068656	-2925.068887	-2925.099180
^3E_g	d_{zz} *and one of* d_{xz} / d_{yz}			
	ΔE (CASSCF)	-2919.421337	-2919.420968	-2919.448387
	ΔE (MRMP2)	-2925.066637	-2925.068232	-2925.085274
$^3B_{2g}$	d_{xy} *and* d_{zz}			
	ΔE (CASSCF)	-2919.38659	-2919.389066	n/a
	ΔE (MRMP2)	-2925.033774	-2925.037525	n/a
Active Space		Calculations		
#occ	Metal	4	4	4
	Ligand	2	1	2
#virt	Ligand	3	2	3
	2nd set of metal	0	4	4
#States weighted:		3	4	4
State and singly occupied metal MOs				
$^3A_{2g}$	d_{xz} and d_{yz}			
		-2919.435019	-2919.499054	-2919.511898
		-2925.073700	-2925.138095	-2925.041339
^3E_g	d_{zz} *and one of* d_{xz} / d_{yz}			
	ΔE (CASSCF)	-2919.434421	-2919.497428	-2919.510554
	ΔE (MRMP2)	-2925.071666	-2925.043603	-2925.04629
$^3B_{2g}$	d_{xy} *and* d_{zz}			
	ΔE (CASSCF)	-2919.399934	-2919.467160	-2919.480602
	ΔE (MRMP2)	-2925.039178	-2925.105606	-2925.016104

Table 3: Absolute energies for the calculations reported in chapter 6, table 3. Low-lying MnPc states calculated at CASSCF/cc-pVDZ and MRMP2/cc-pVDZ. States are characterized by their D_{4h} symmetry labels and by which metal orbital is doubly occupied.

Active Space		Calculations		
#occ	Metal	4	4	4
	Ligand	1	2	2
#virt	Ligand	2	2	2
	2nd set of metal	0	0	0
#States weighted:		4	4	15
State and doubly-occupied metal MOs				
$^4B_{2g}$	d_{xy}			
	CASSCF	-2806.793371	-2806.796389	-2806.795230
	MRMP2	-2812.39026	-2812.381213	-2812.383736
^4E_g	d_{xz} / d_{yz}			
	ΔE (CASSCF)	-2806.774617	-2806.778155	-2806.774536
	ΔE (MRMP2)	-2812.377236	-2812.369550	-2812.373938
$^4A_{1g}$	d_{z^2}			
	ΔE (CASSCF)	-2806.773102	-2806.776856	-2806.773519
	ΔE (MRMP2)	-2812.377614	-2812.369840	-2812.372633
Active Space		Calculations		
#occ	Metal	4	4	4
	Ligand	2	1	2
#virt	Ligand	2	2	3
	2nd set of metal	0	4	4
#States weighted:		6	4	4
State and doubly-occupied metal MOs				
$^4B_{2g}$	d_{xy}			
	CASSCF	-2806.797966	-2806.845041	-2806.857937
	MRMP2	-2812.381007	-2812.361303	-2812.365111
^4E_g	d_{xz} / d_{yz}			
	ΔE (CASSCF)	-2806.775524	-2806.828466	-2806.841830
	ΔE (MRMP2)	-2812.371097	-2812.347837	-2812.352028
$^4A_{1g}$	d_{z^2}			
	ΔE (CASSCF)	-2806.774018	-2806.830201	-2806.843145
	ΔE (MRMP2)	-2812.370458	-2812.347829	-2812.352081